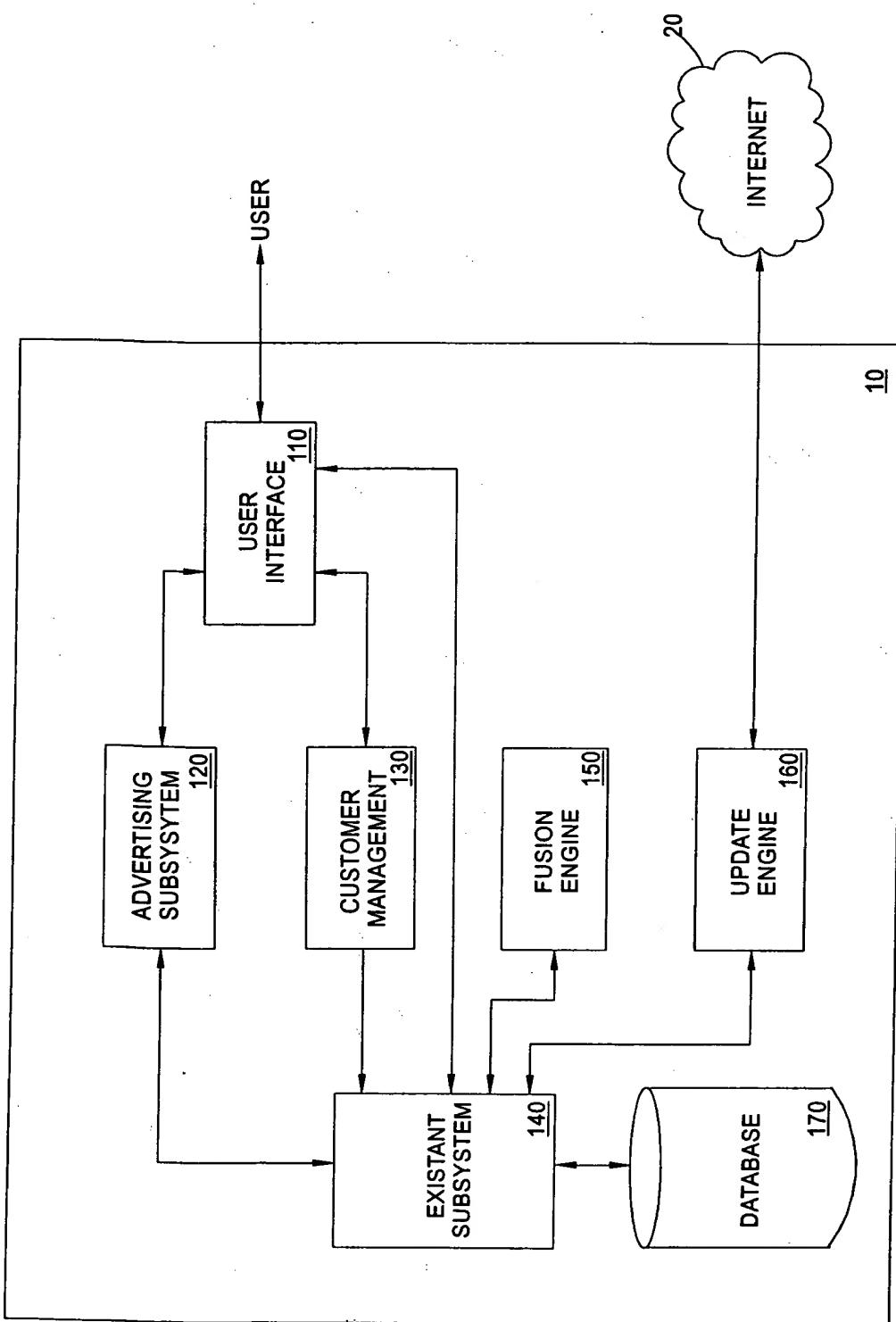


FIG. 1

FIG. 2



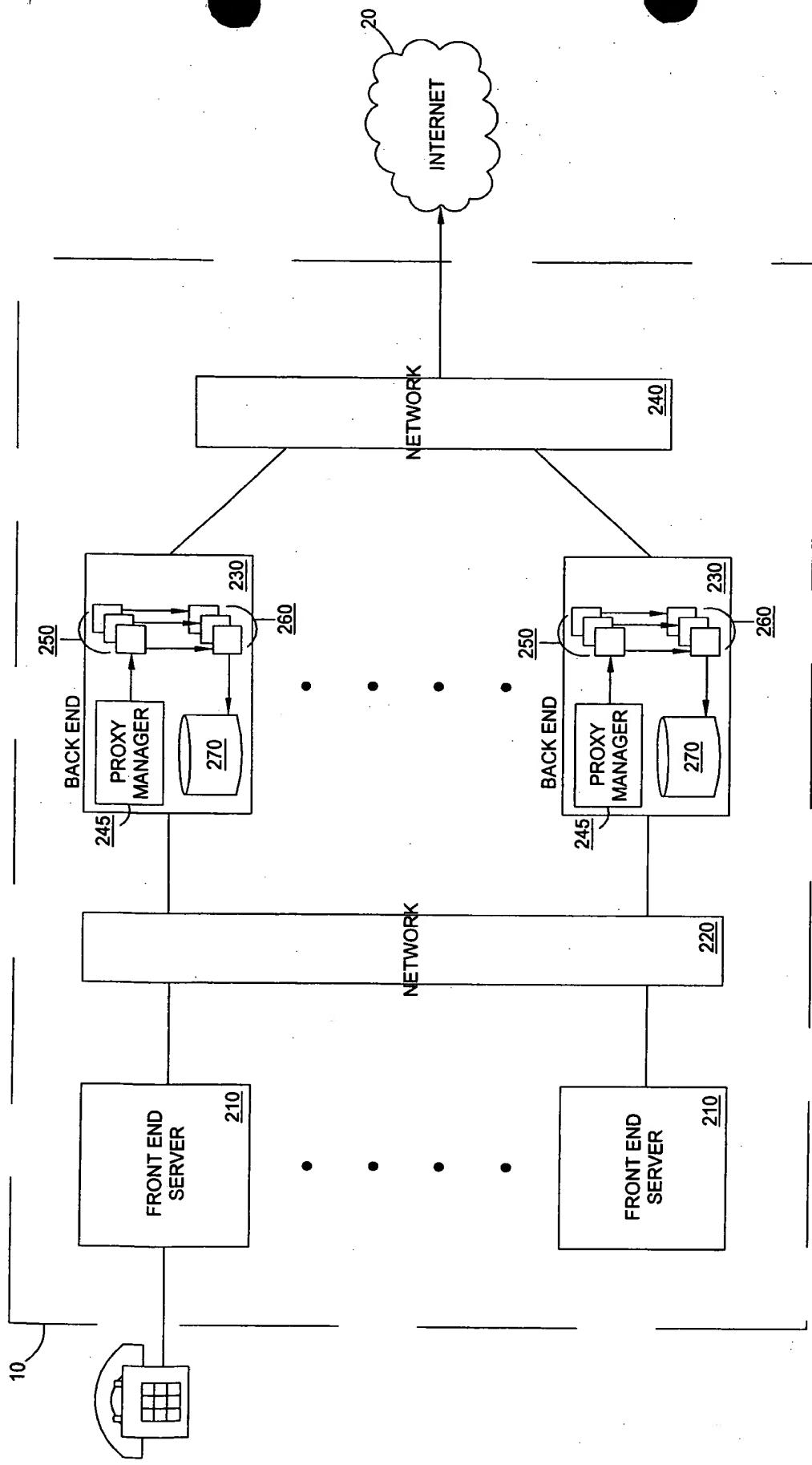


FIG. 3

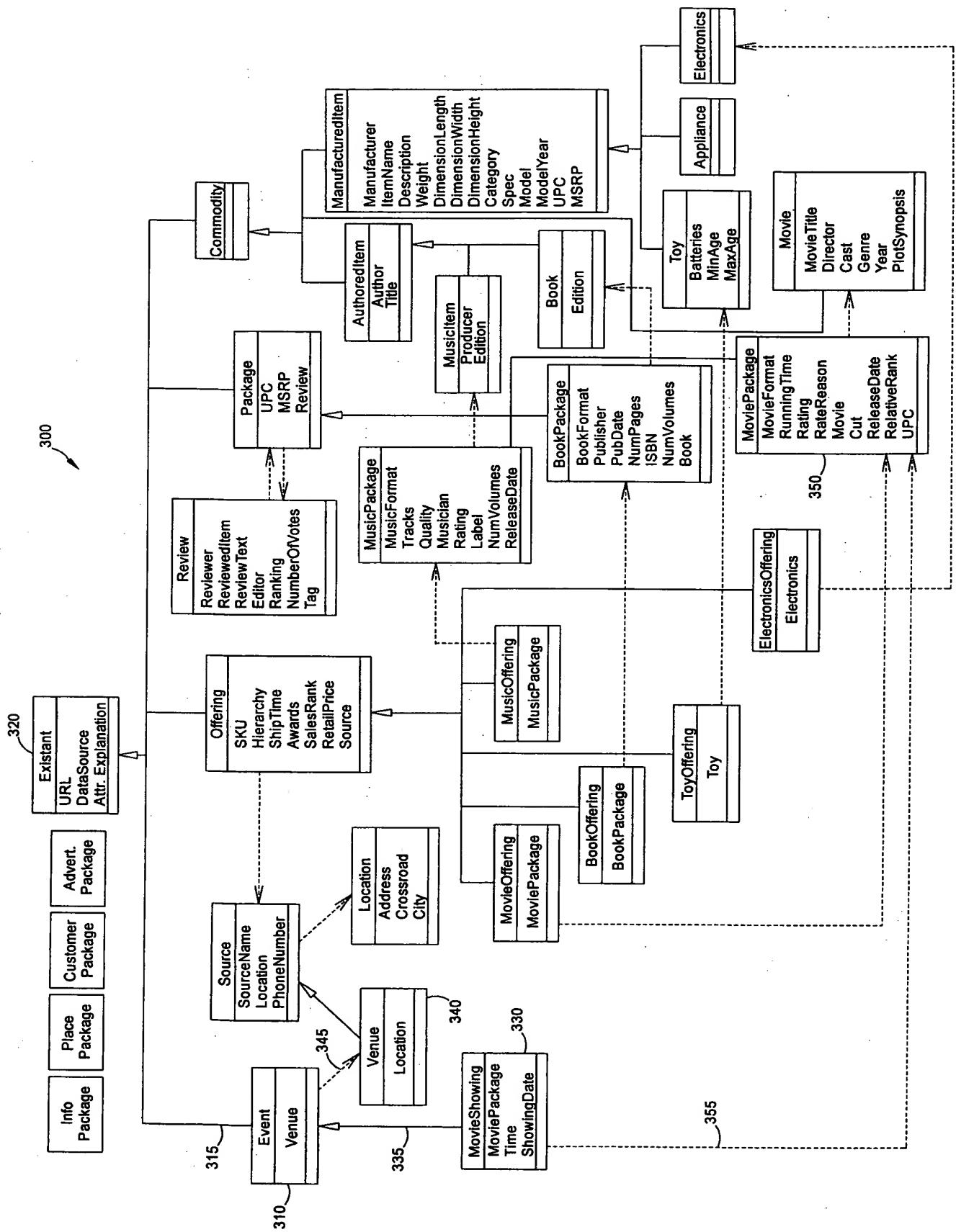
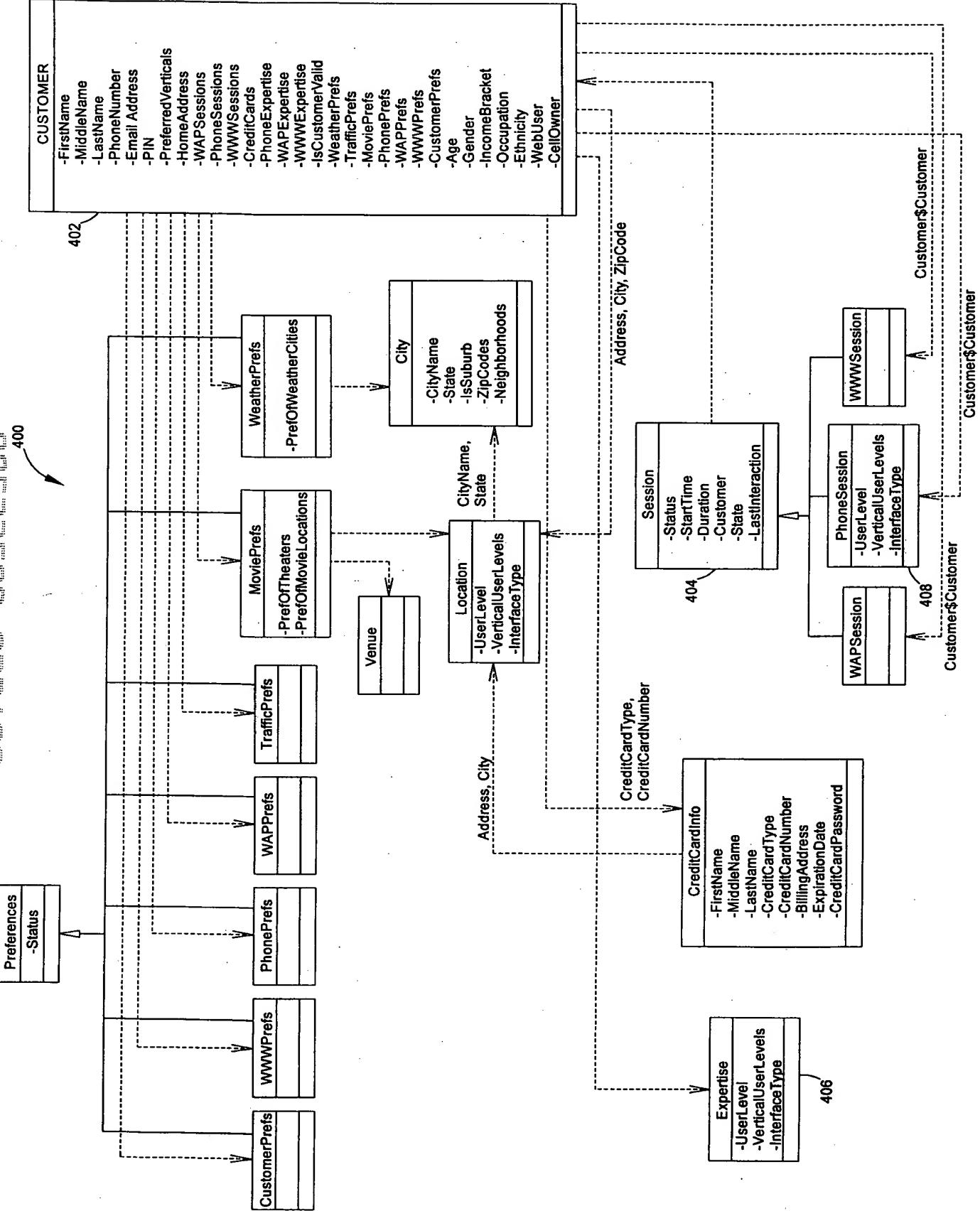


FIG. 4

FIG. 5



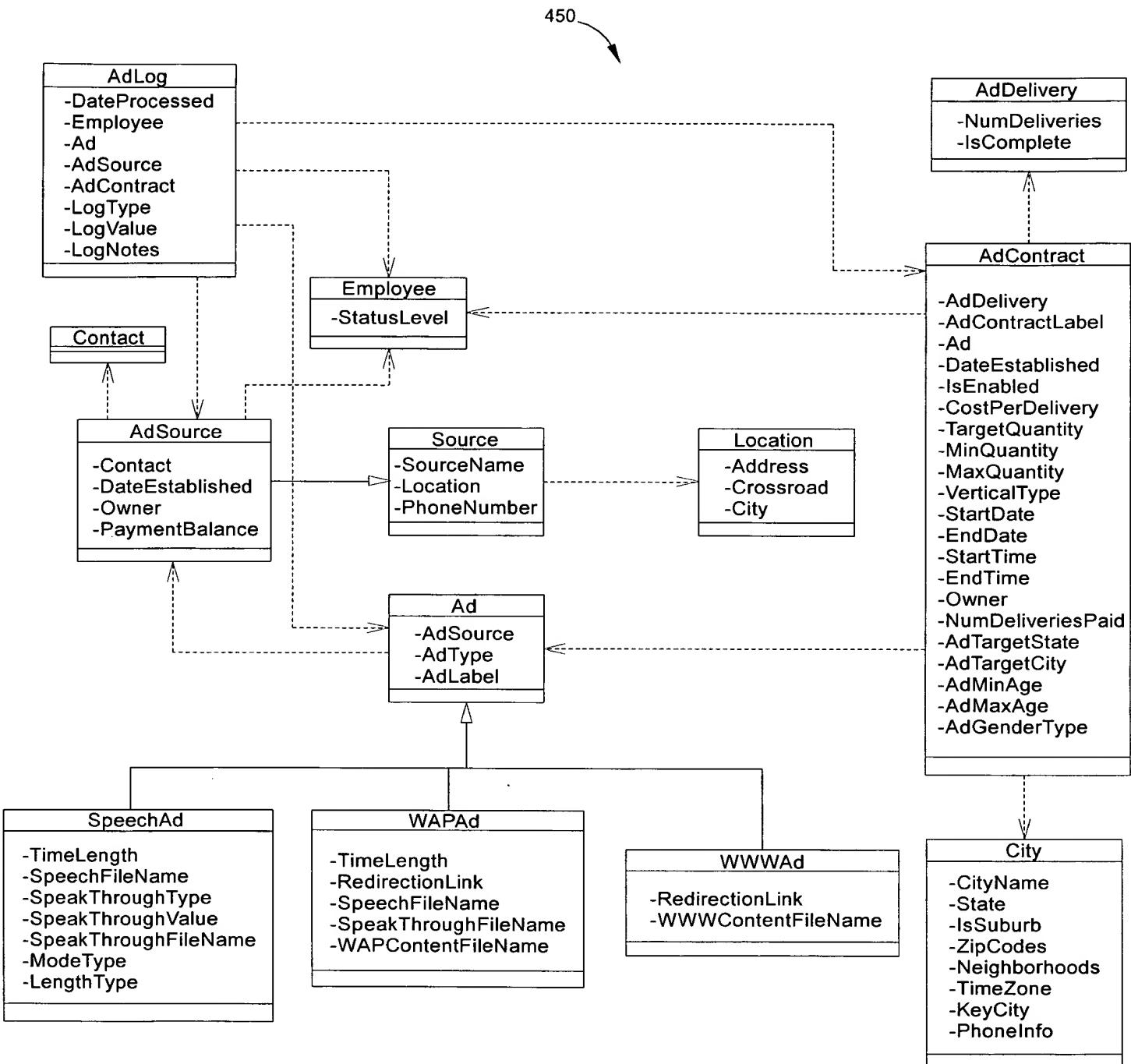


FIG. 6

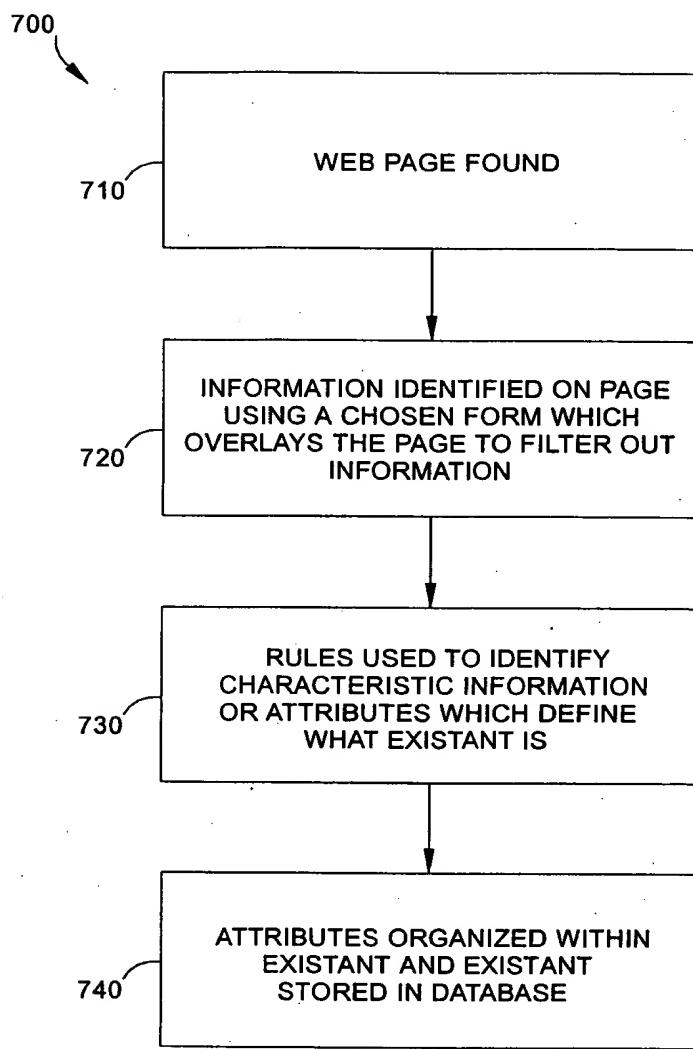
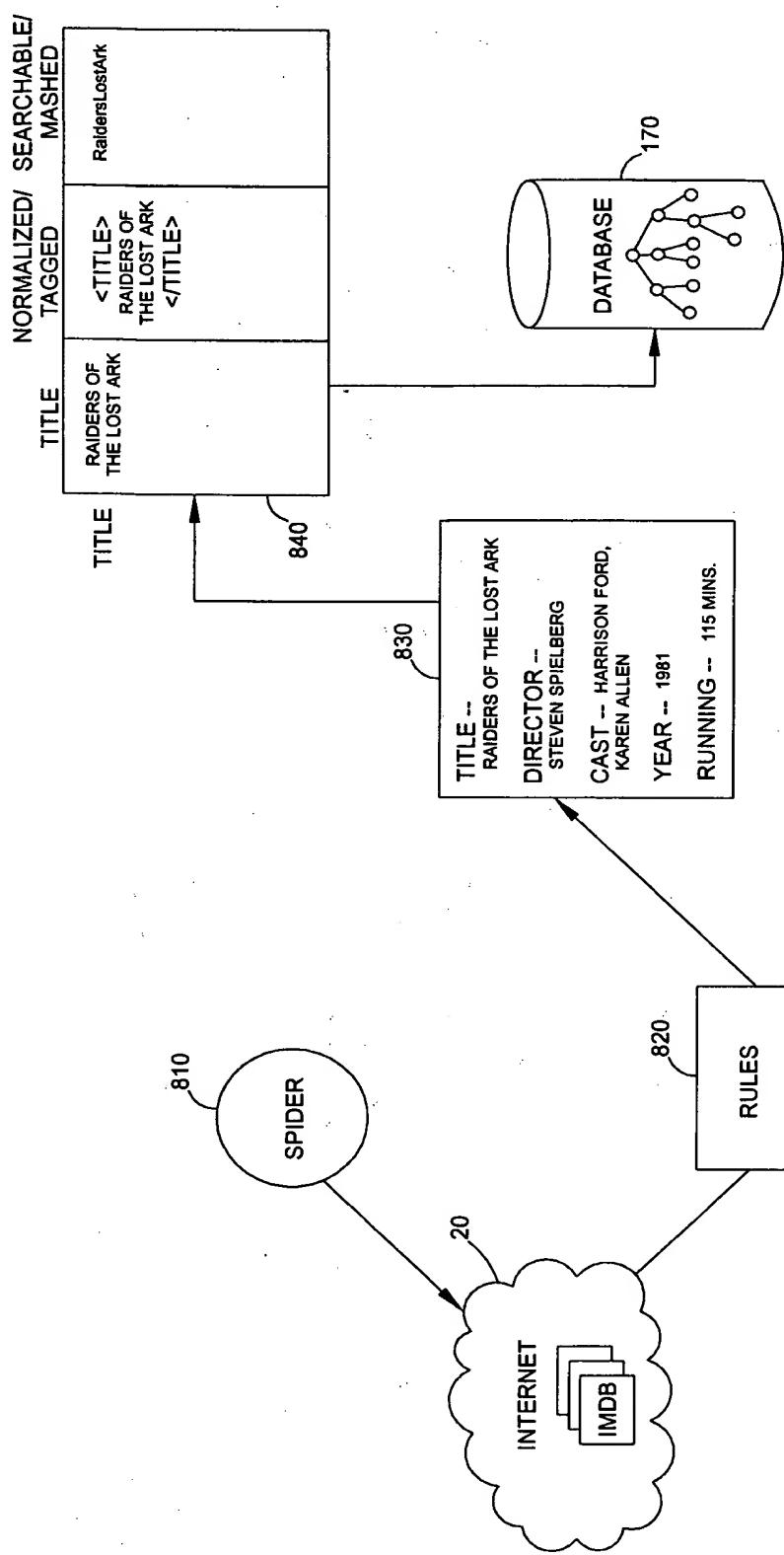


FIG. 7

FIG. 8



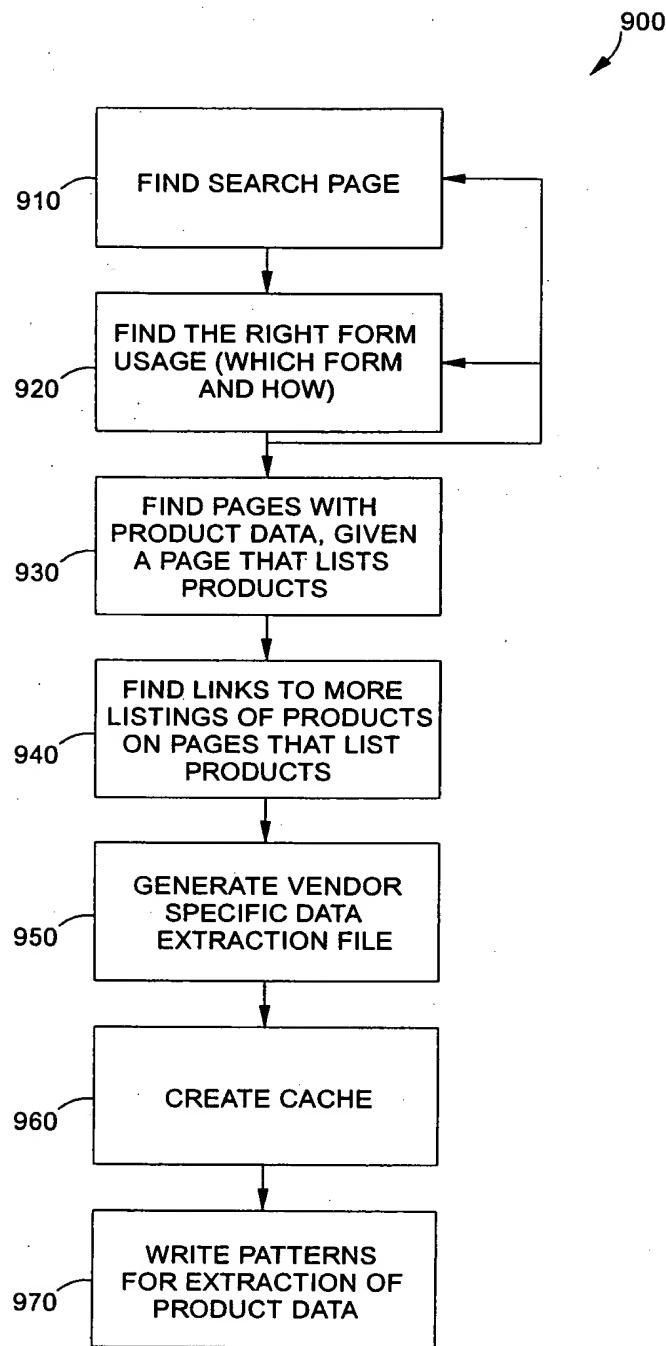


FIG. 9

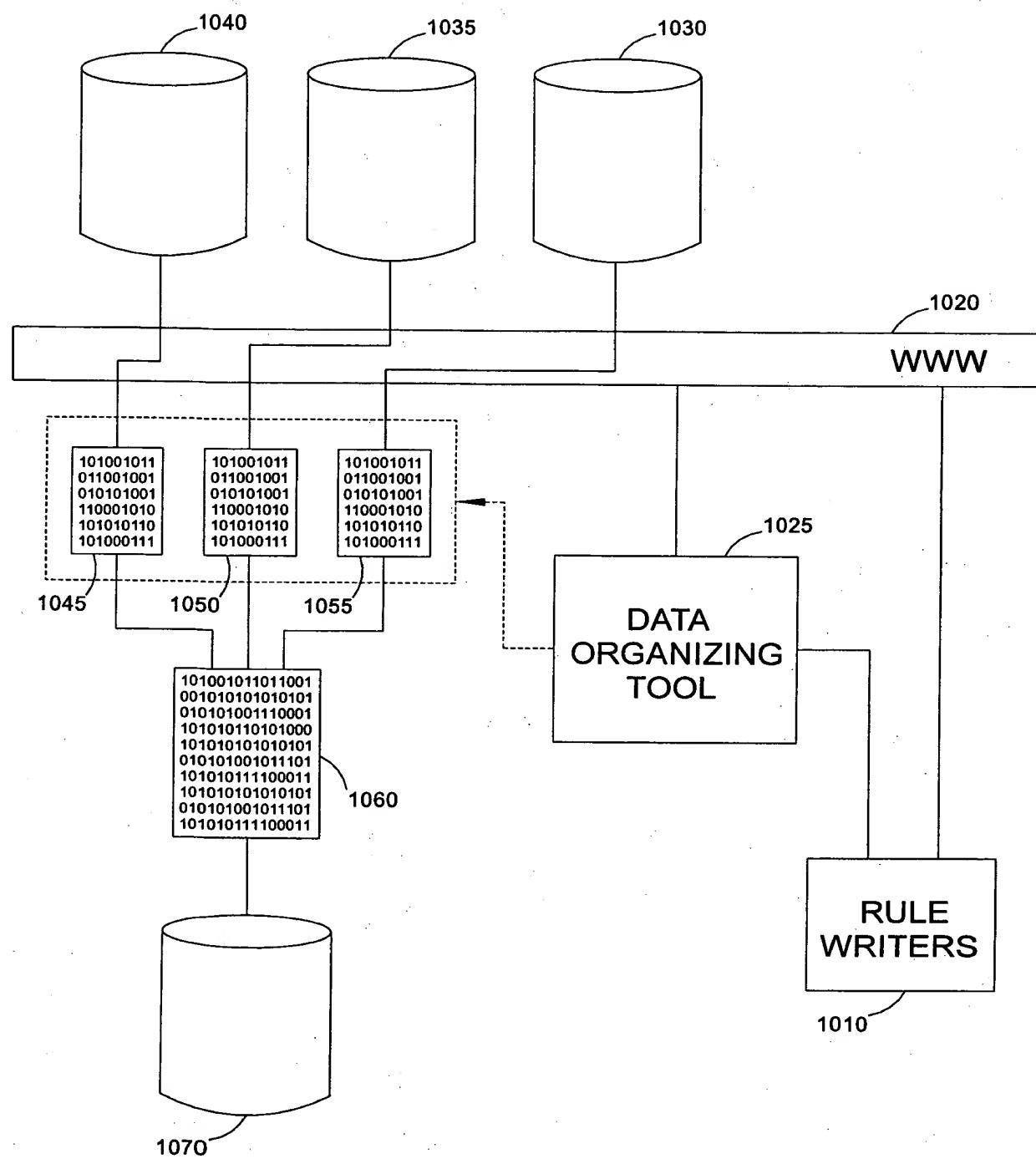


FIG. 10

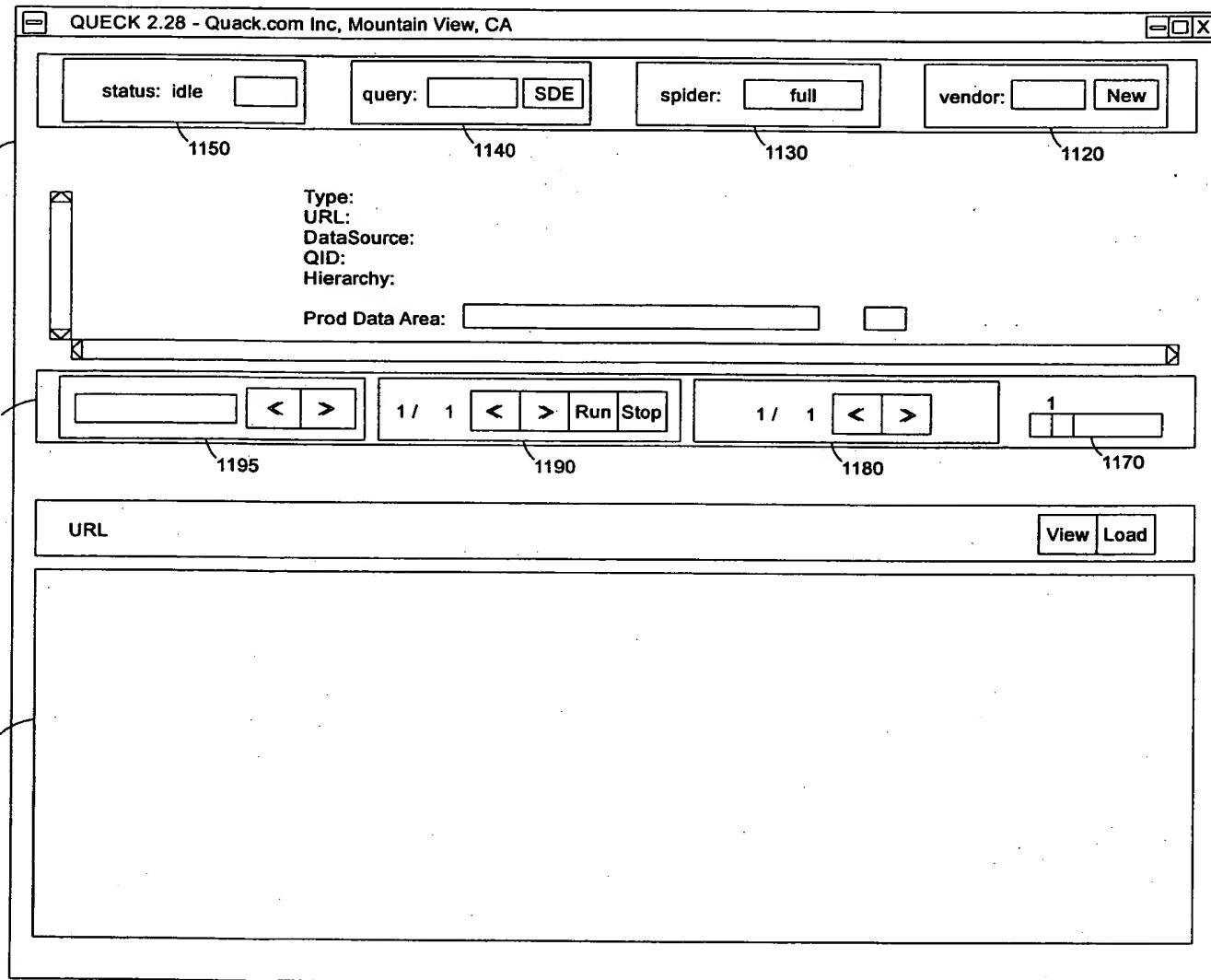


FIG. 11

1200

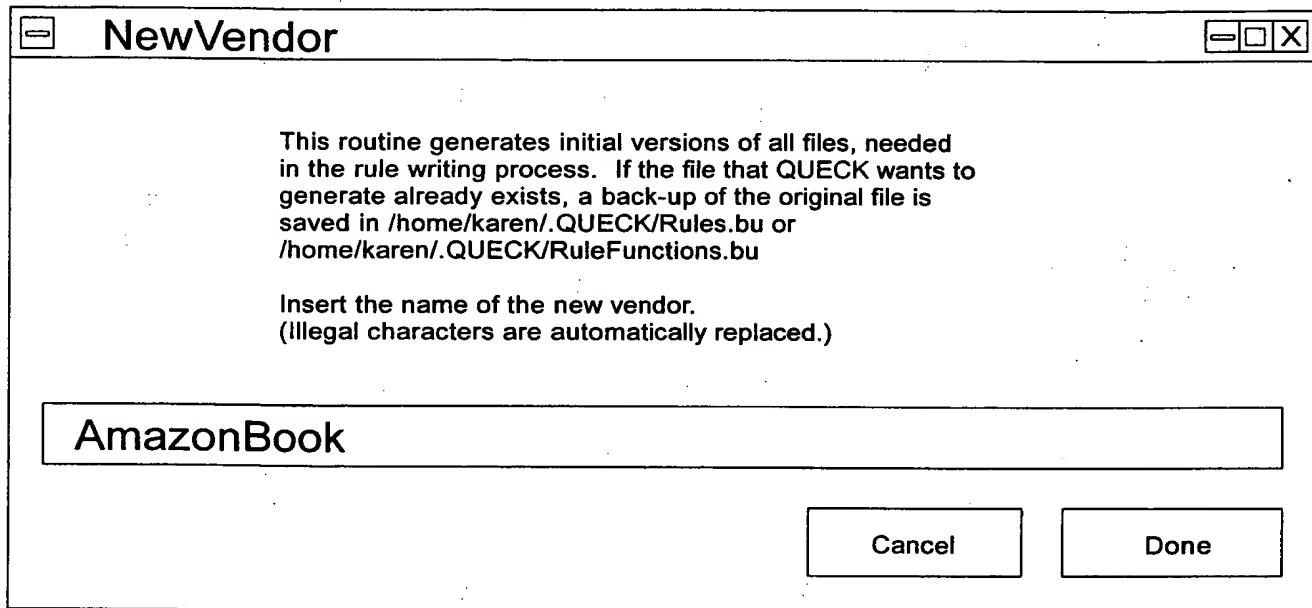


FIG. 12

1300

NewVendor

This routine generates initial versions of all files, needed in the rule writing process. If the file that QUECK wants to generate already exists, a back-up of the original file is saved in /home/karen/.QUECK/Rules.bu or /home/karen/.QUECK/RuleFunctions.bu

Insert the name of the new vendor.
(Illegal characters are automatically replaced.)

AmazonBook

Insert the URL of the AmazonBook main page.

http://www.amazonbook.com

Cancel **Done**

FIG. 13

1400

Search Data Editor

1410 ISBN:

1420 UPC:

1430 ◆ ◇ ◇ ◇ ◇
◇ ◇ ◇ ◇ ◇

Clear **Reset** **Save** **Done**

FIG. 14

1500

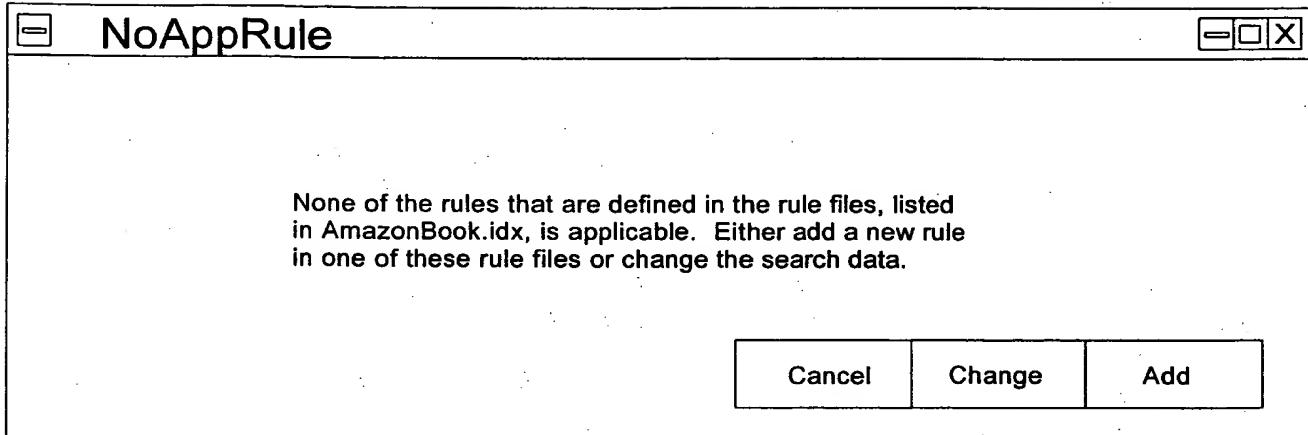


FIG. 15

1600

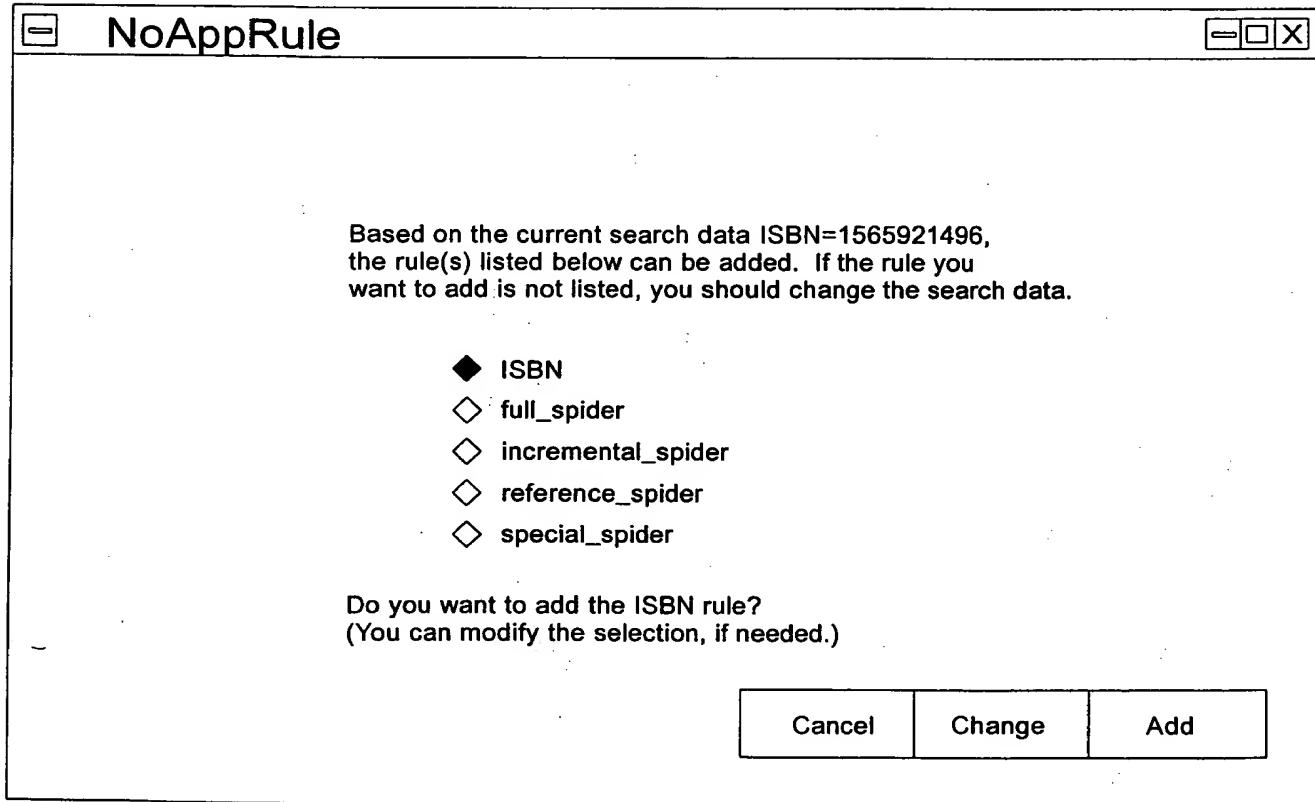


FIG. 16

1700

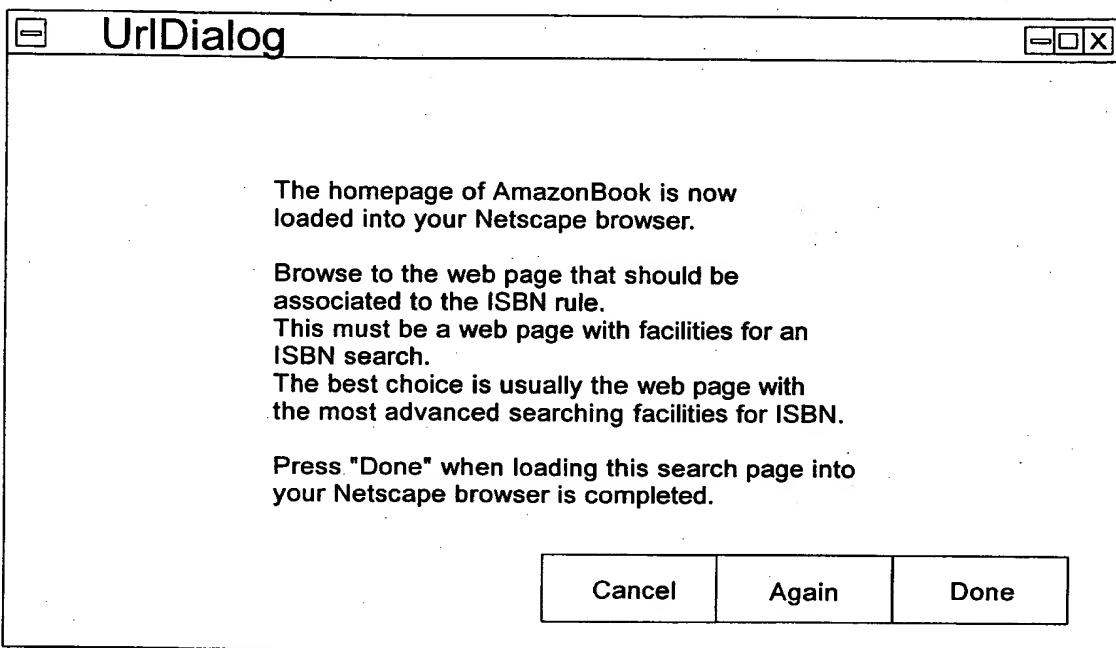


FIG. 17

1800

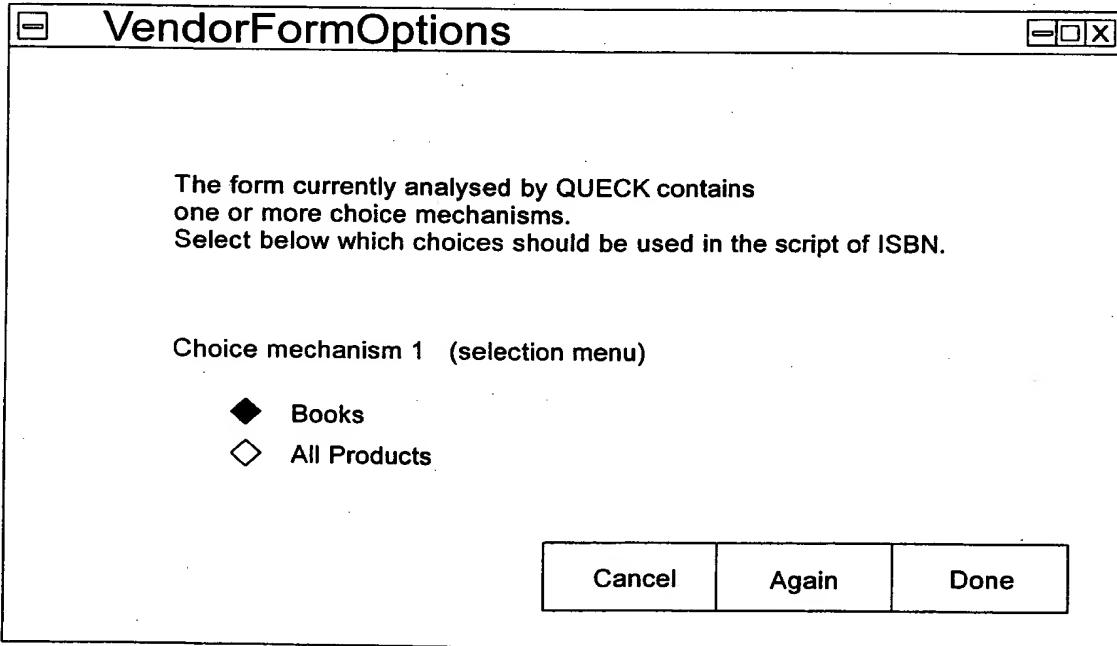


FIG. 18

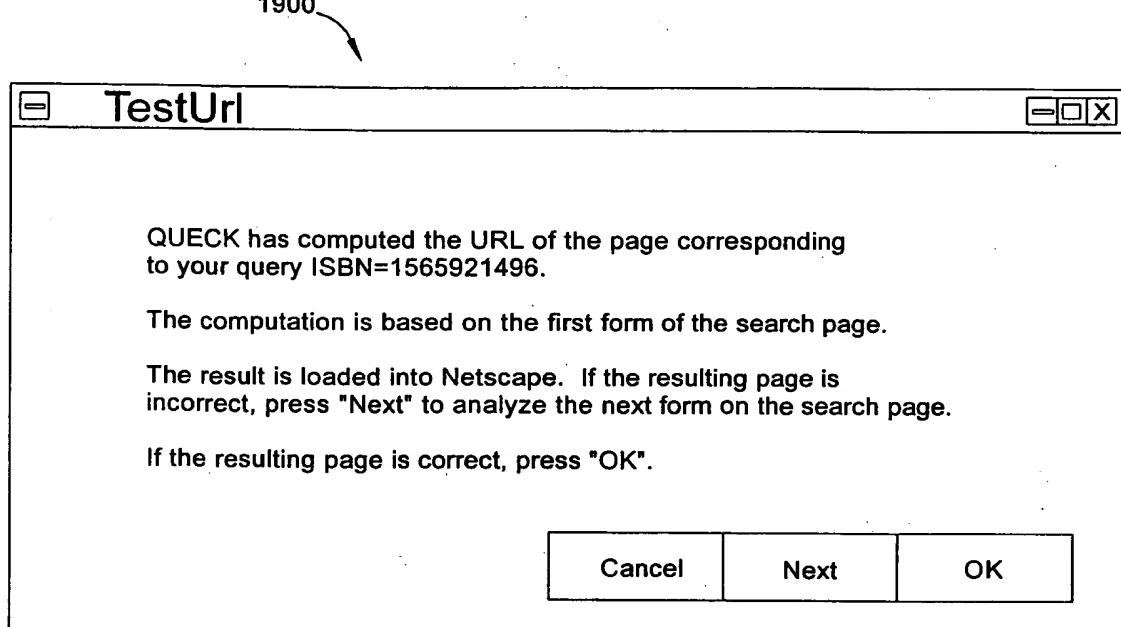


FIG. 19

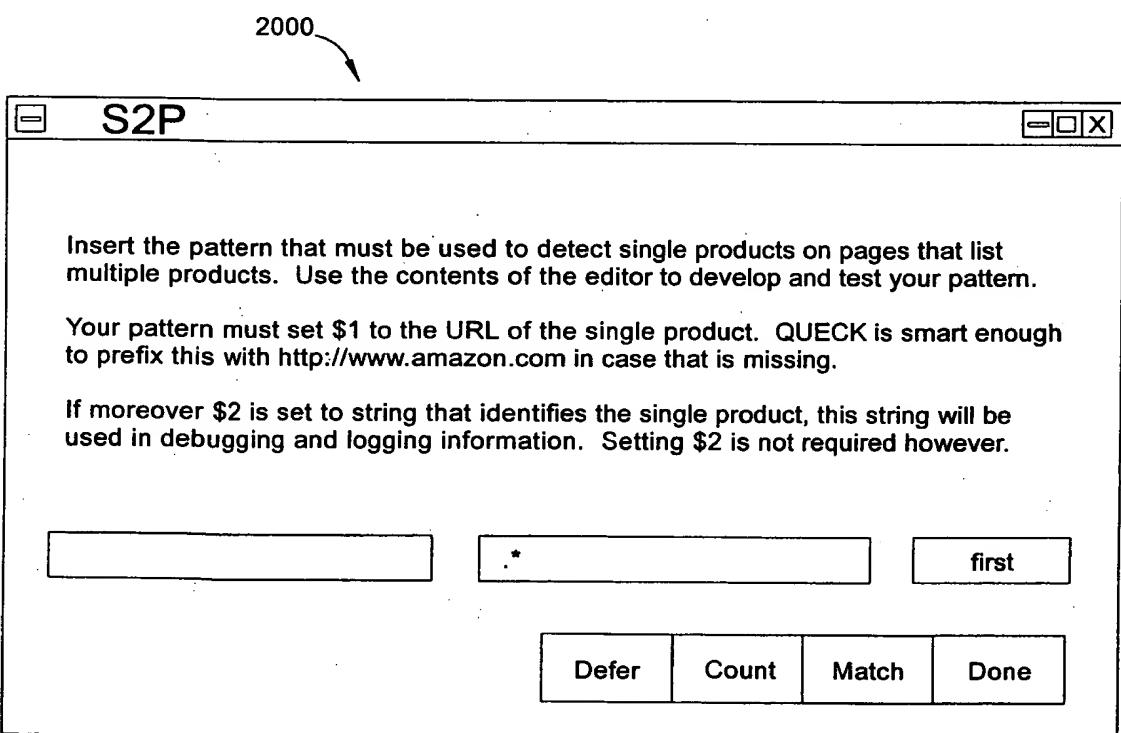


FIG. 20

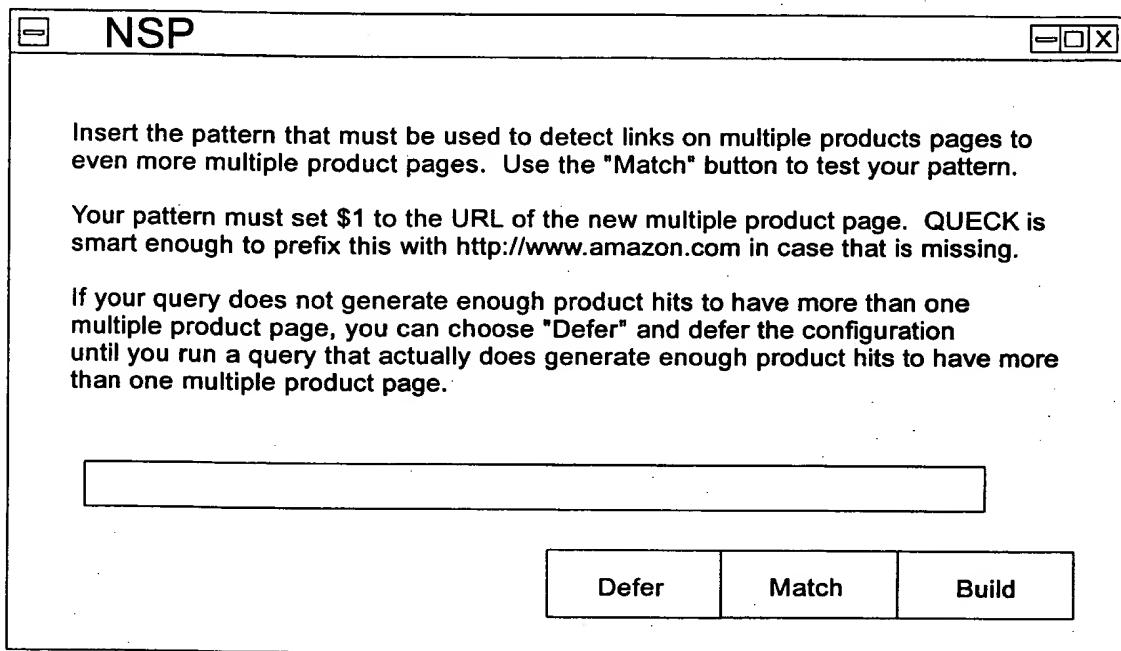


FIG. 21

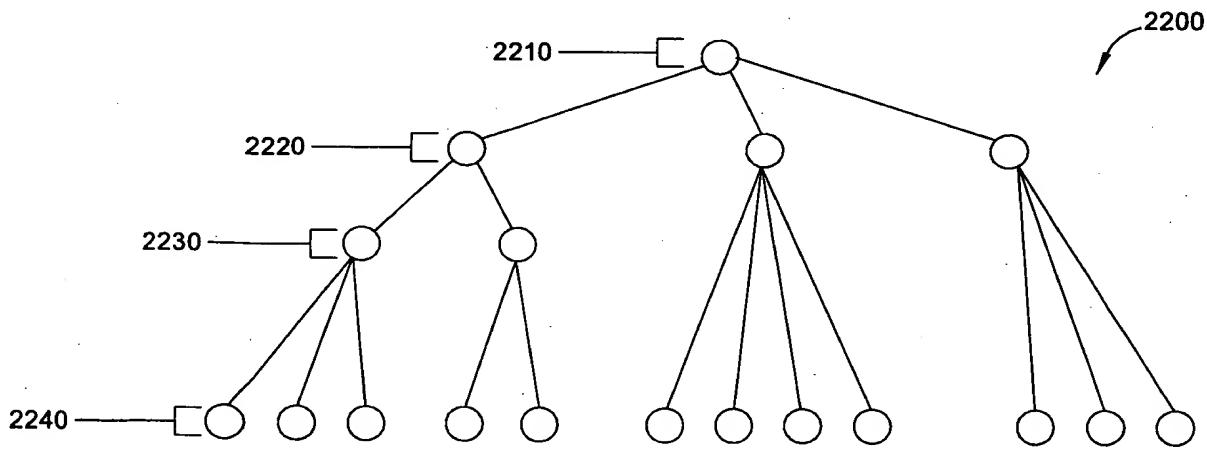


FIG. 22

2300

SpiderSubr

Insert here the URL of the page, currently loaded into Netscape. This is the page associated to the full_spider rule.

Next, set "SpiderDepth" to the maximum number of links that has to be followed from the top of the hierarchy to the actual product pages. Note that in some cases this number depends on the branch you follow. Setting "SpiderDepth" too low creates a spider that misses products that are nested too deep in the hierarchy. Setting "SpiderDepth" too high leads to a decrease in performance.

SpiderDepth UpperBound

1 0

Done

FIG. 23

2400

SpiderSubr

<http://www.amazon.com/exec/obidos/subst/home/home.html/002-5797861-2625002>

The spider you specified is a level - 1 spider.
This means that your spider has the following form:

level - 0: The top page (accessed via the URL above)
level - 1: The single product pages to be spidered

Insert below the pattern used to detect level - 1 pages on the top page.

Your pattern must set \$1 to the URLs of the child pages. QUECK is smart enough to prefix this URL with <http://www.amazon.com> in case it is missing. If your pattern also sets \$2, that value will be used in the hierarchy attributes.

first 1st Level *

Cancel Count Match Build

FIG. 24

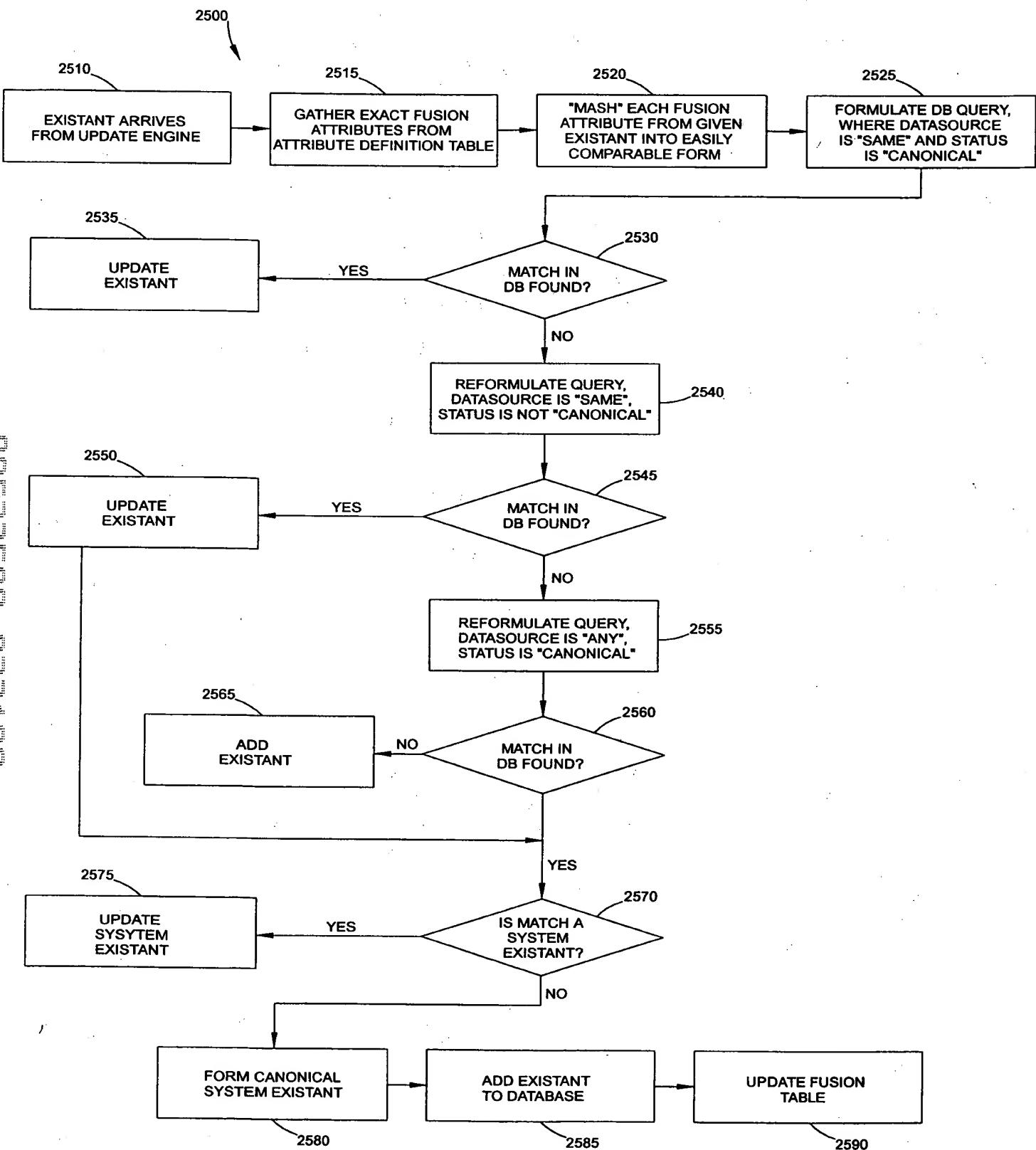


FIG. 25

00000000000000000000000000000000

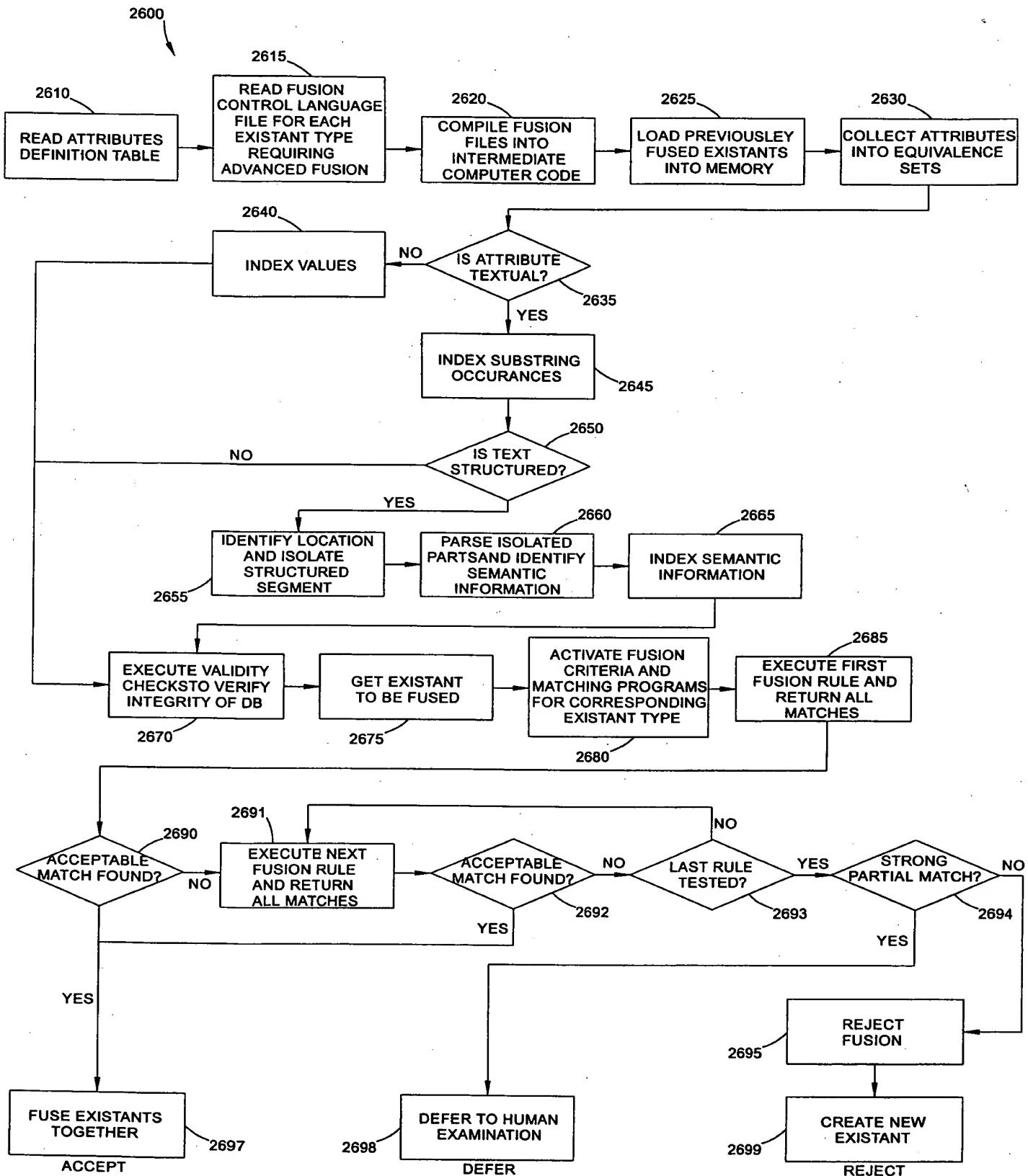


FIG. 26

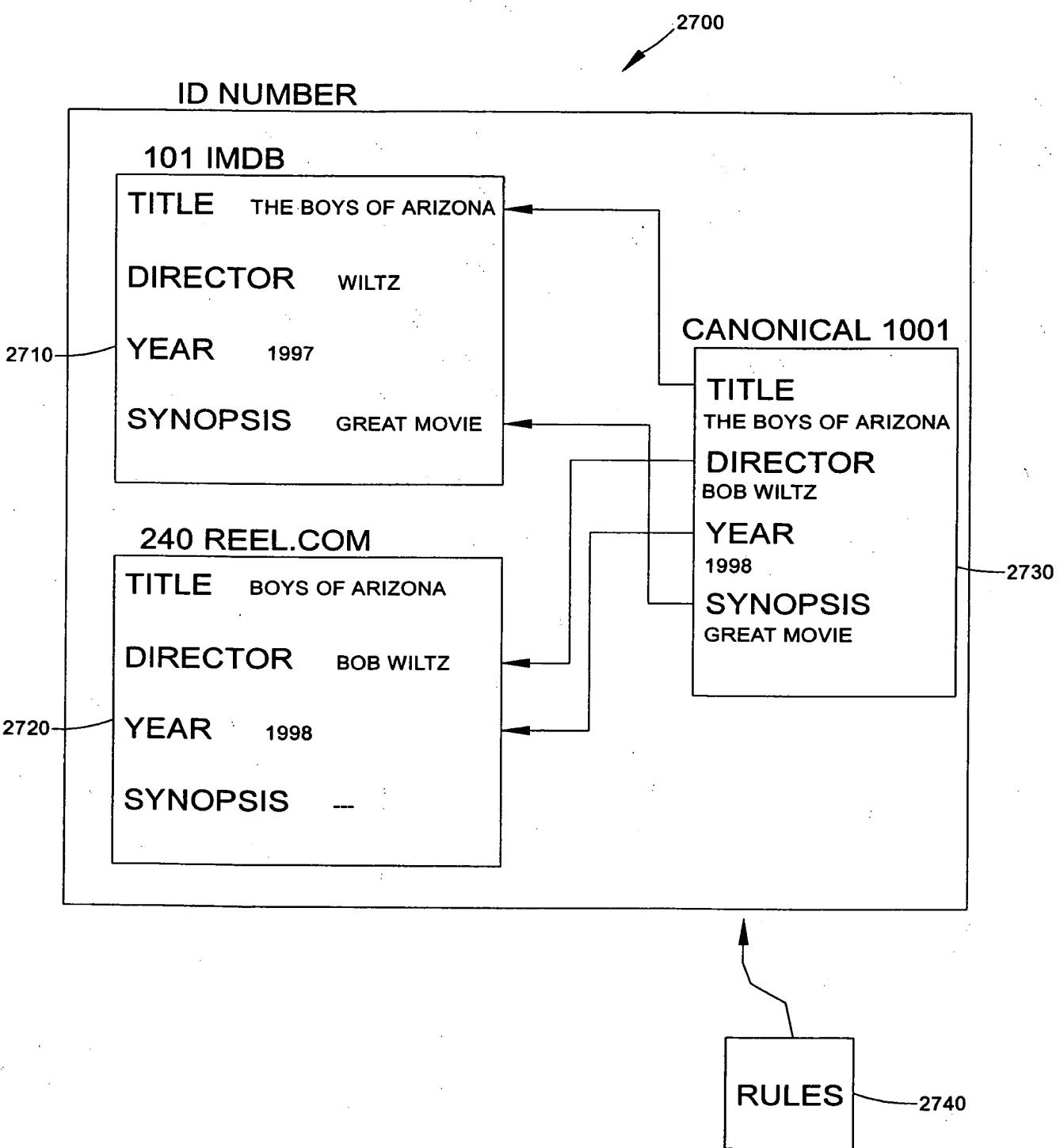


FIG. 27

FIG. 28

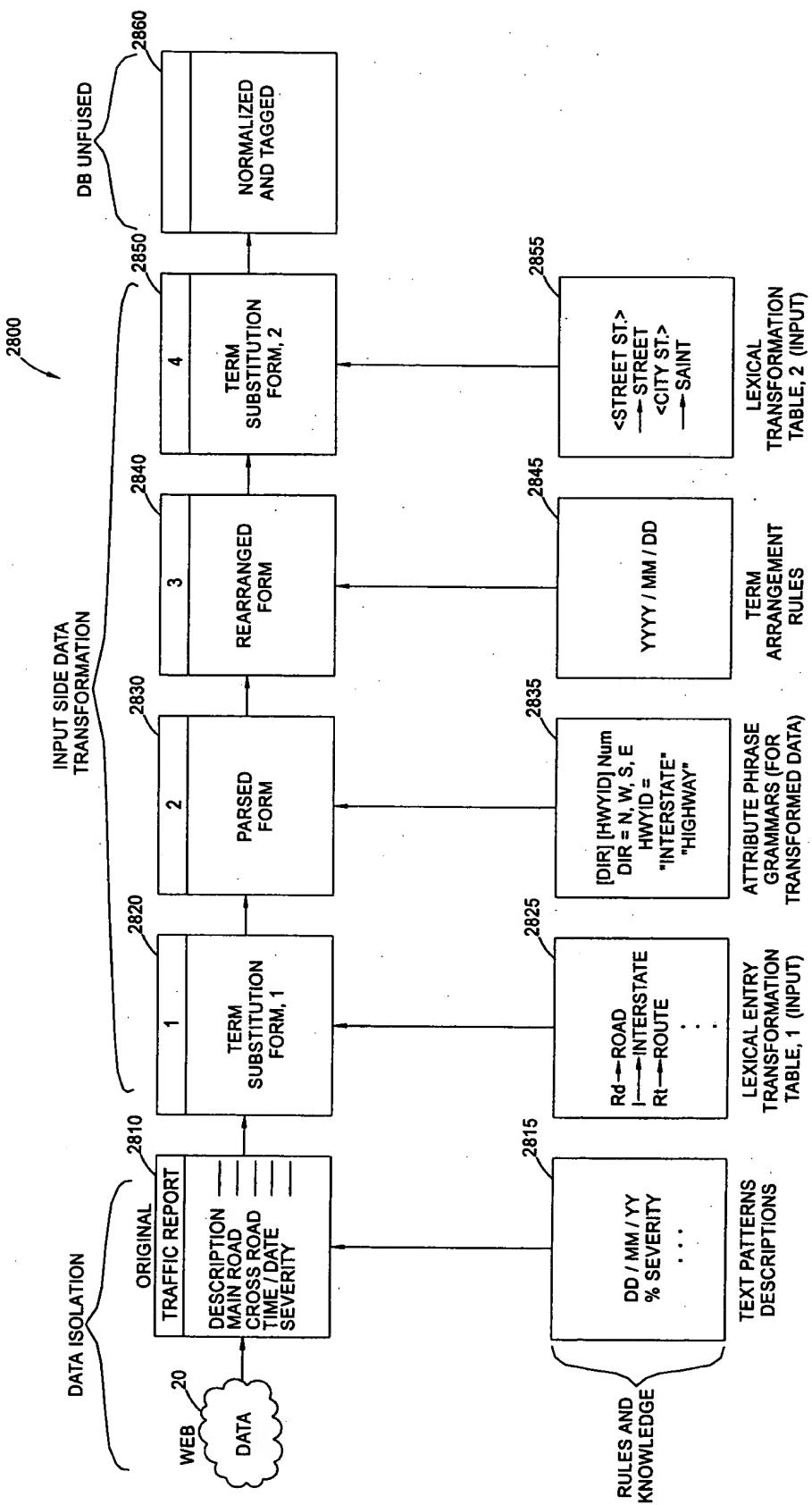
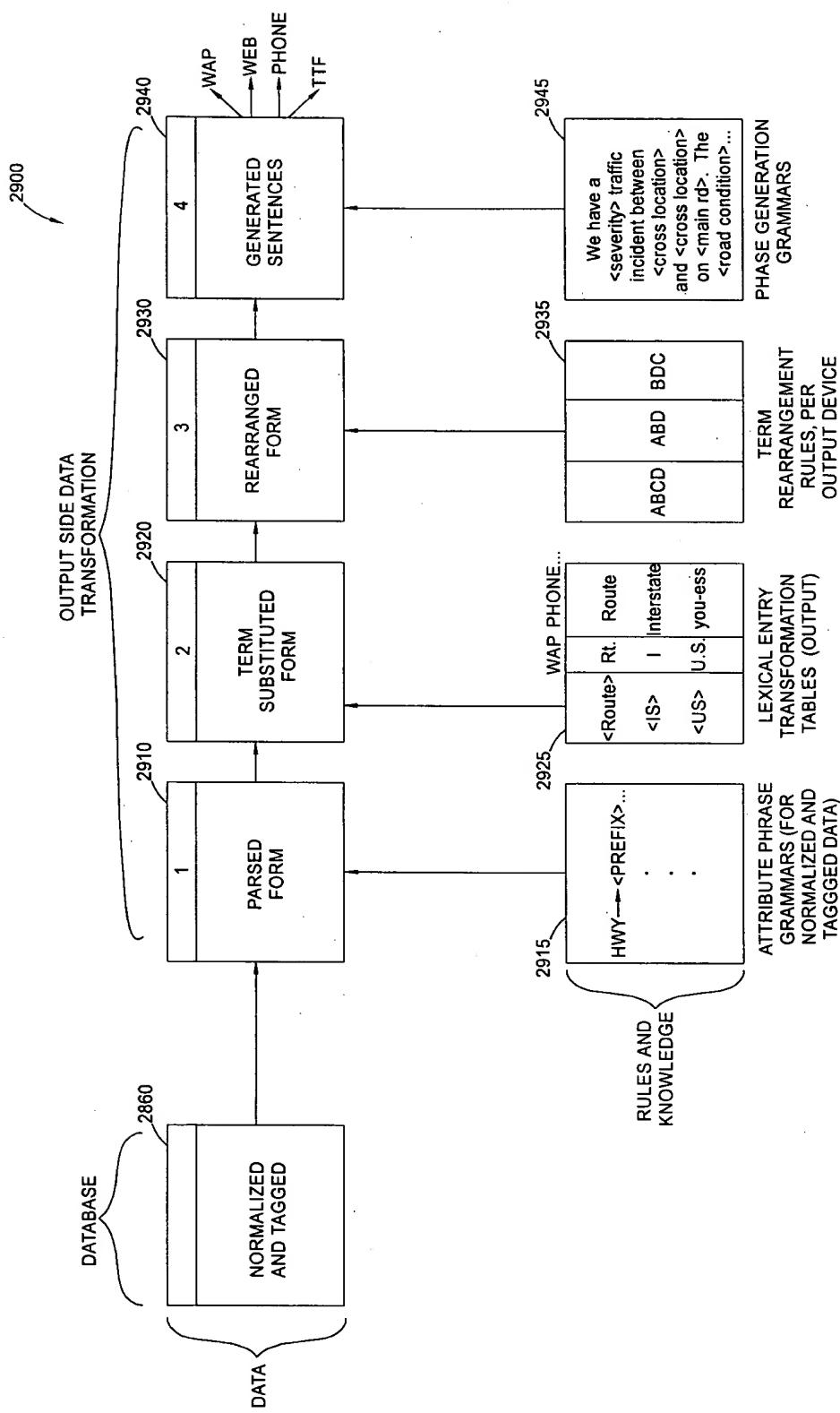


FIG. 29



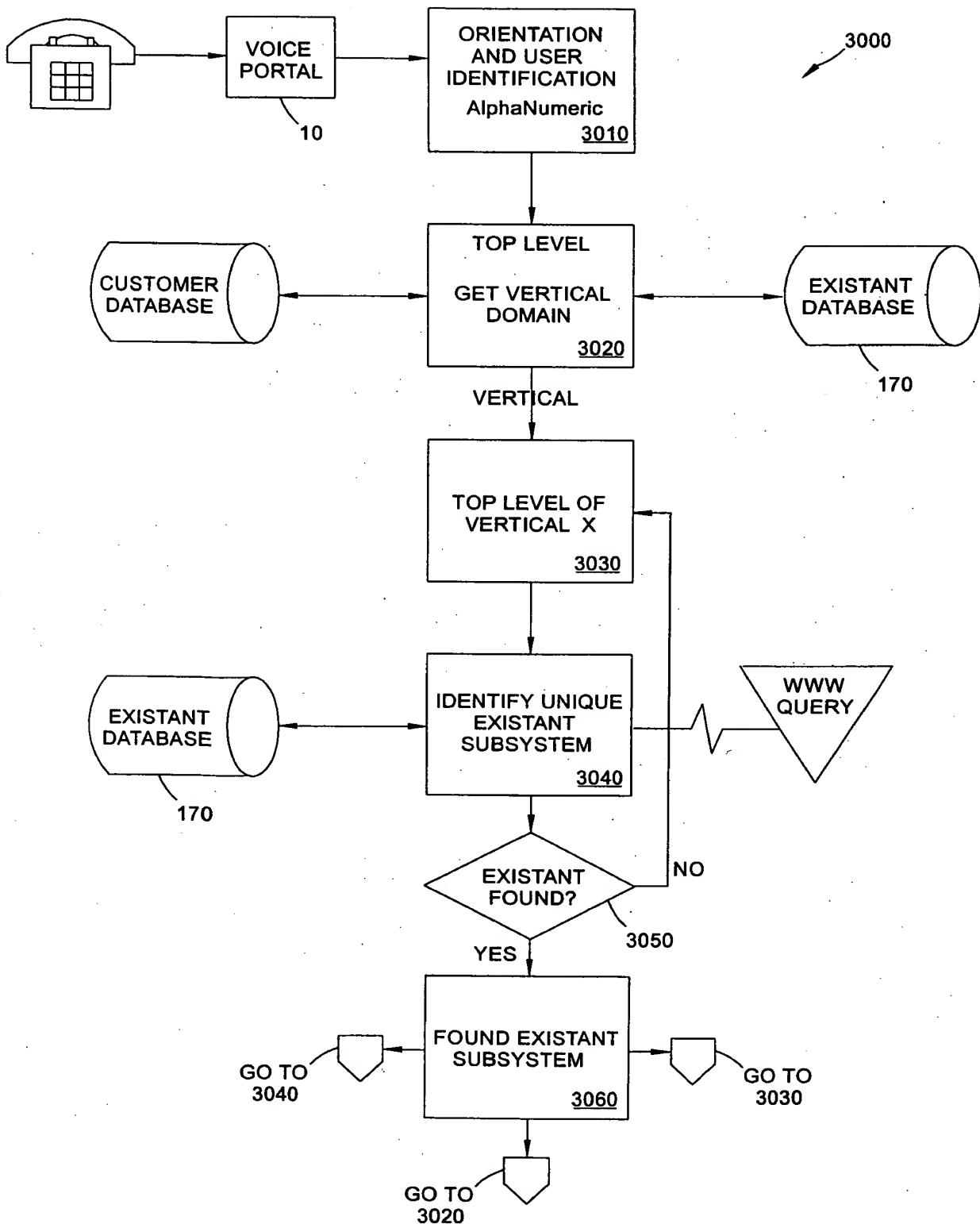


FIG. 30

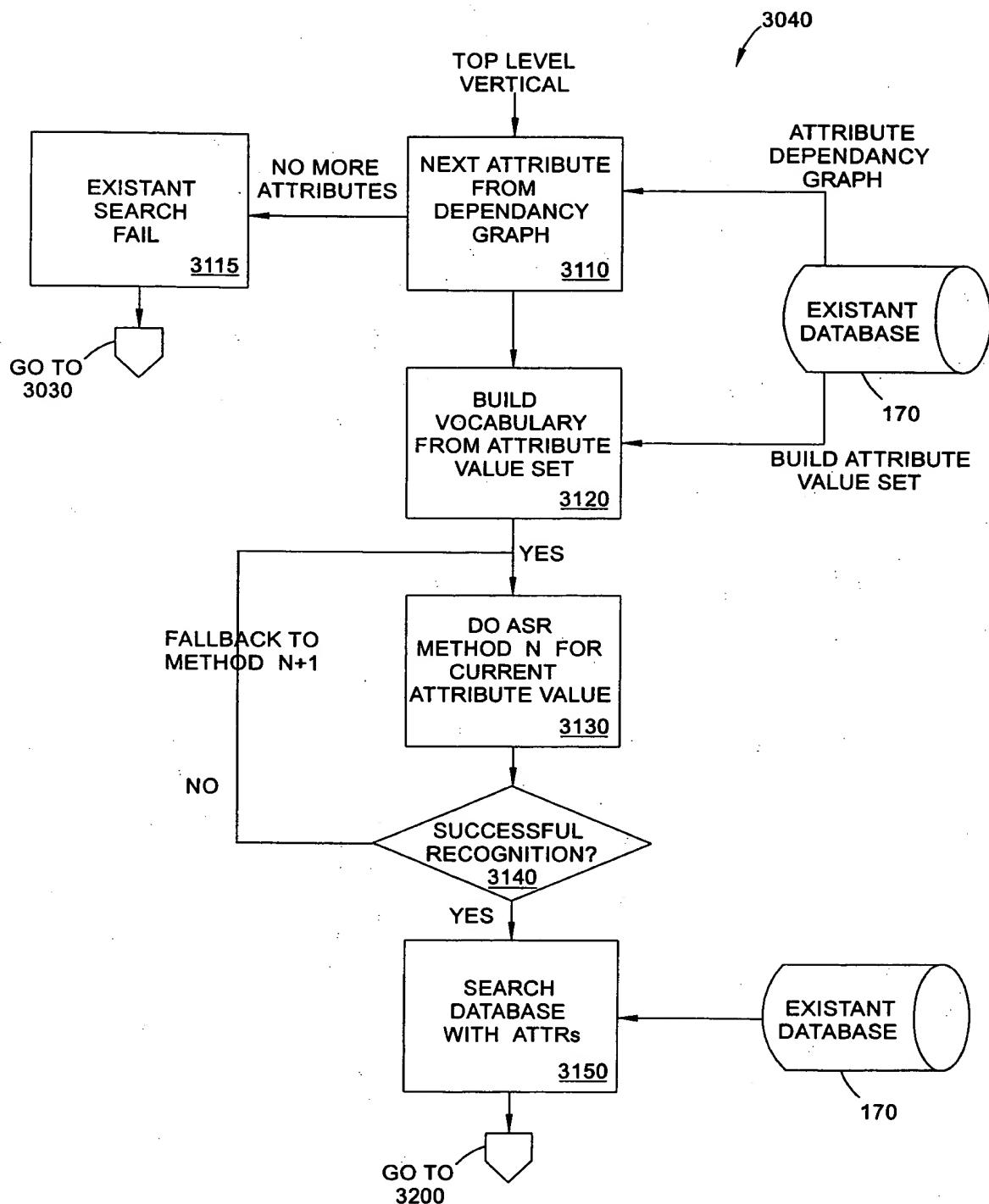


FIG. 31

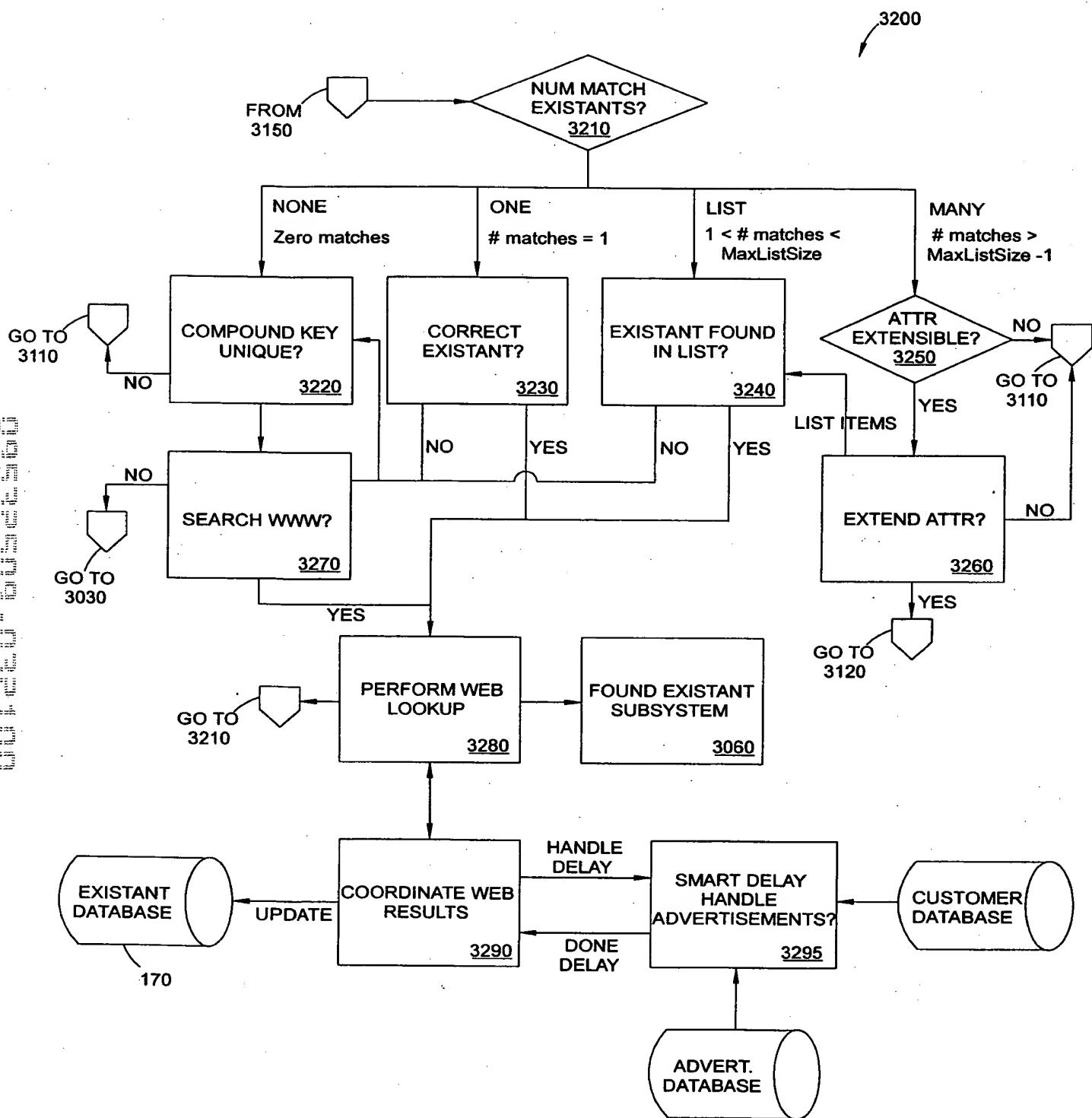


FIG. 32

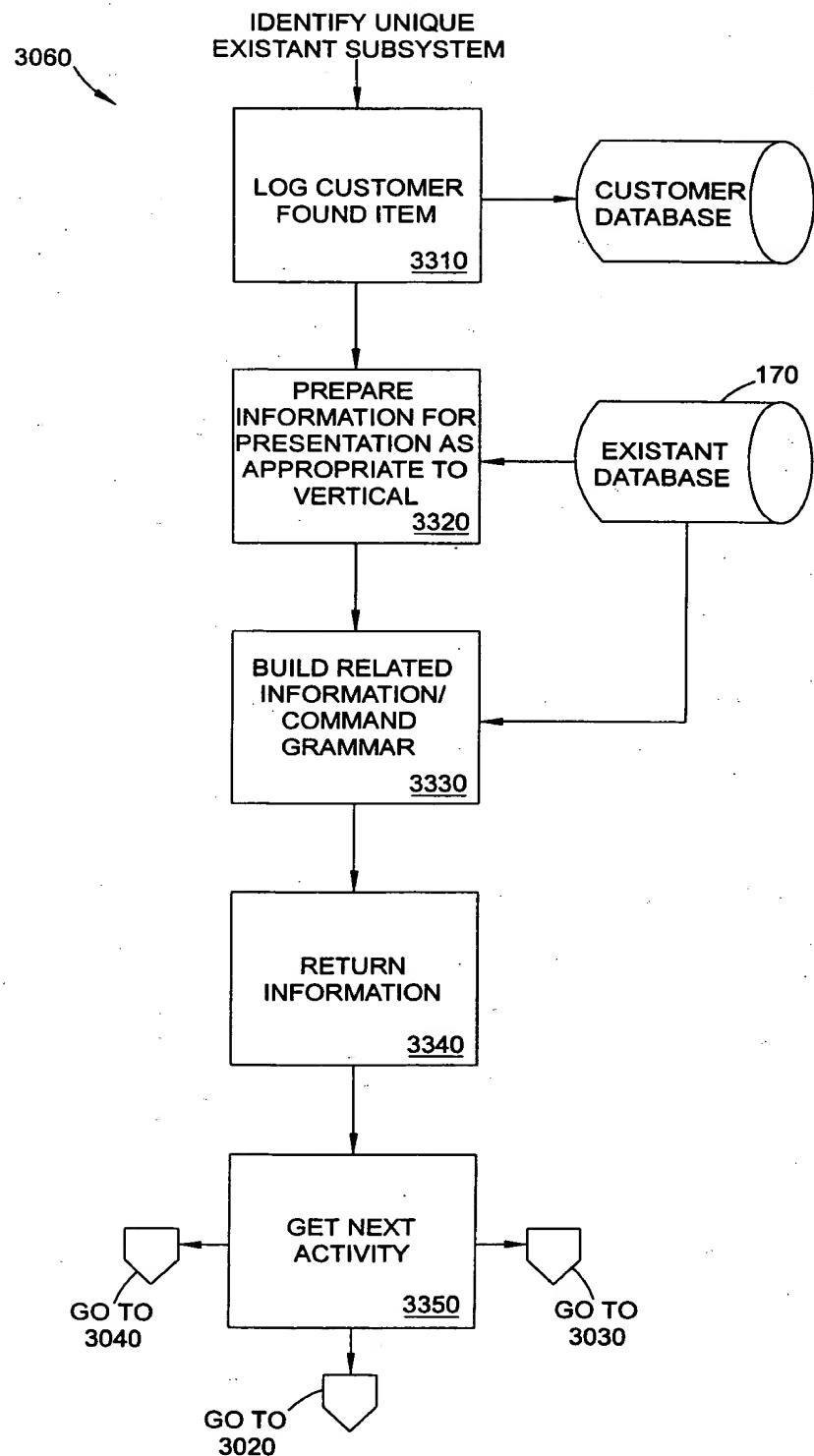


FIG. 33

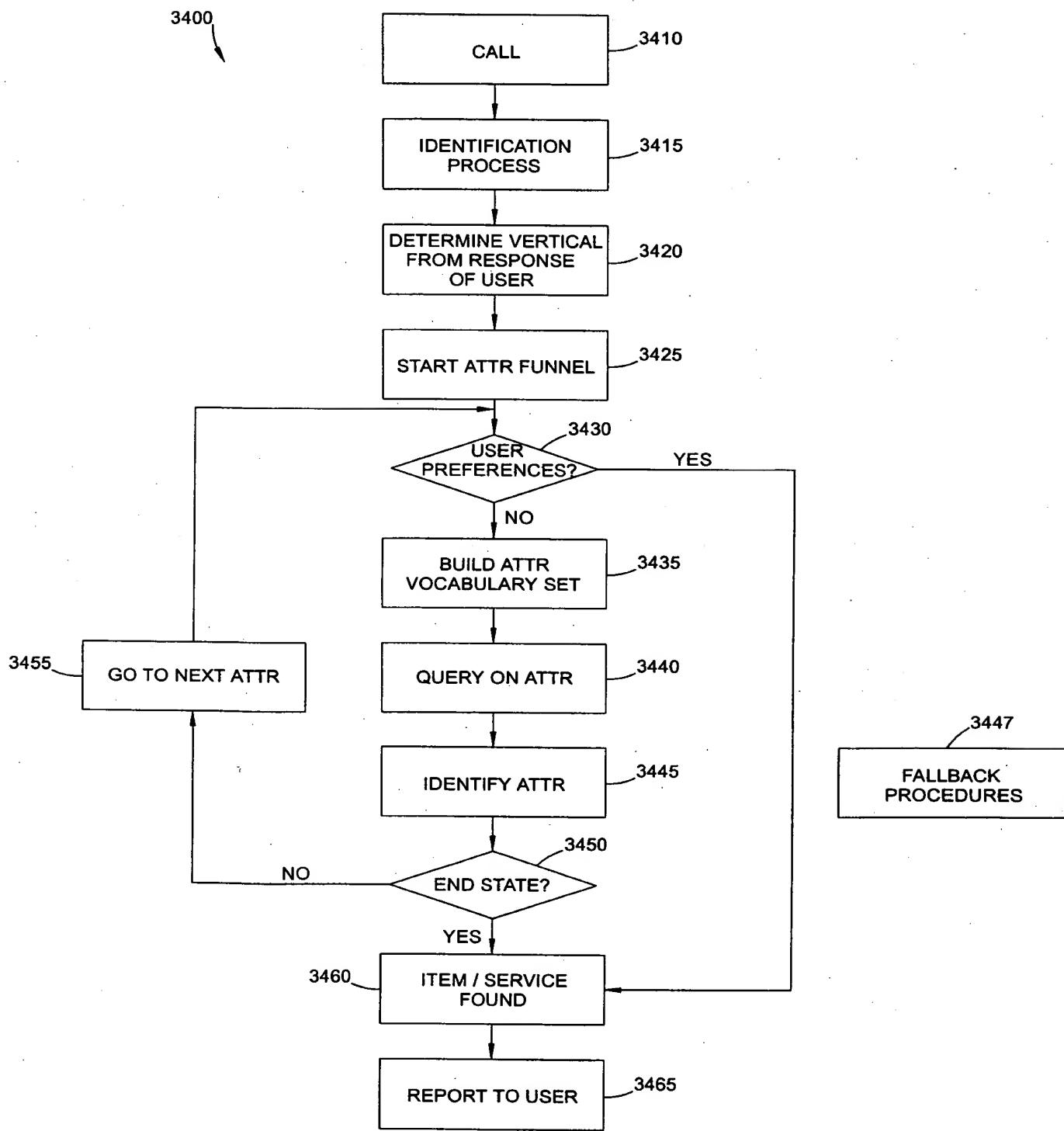


FIG. 34

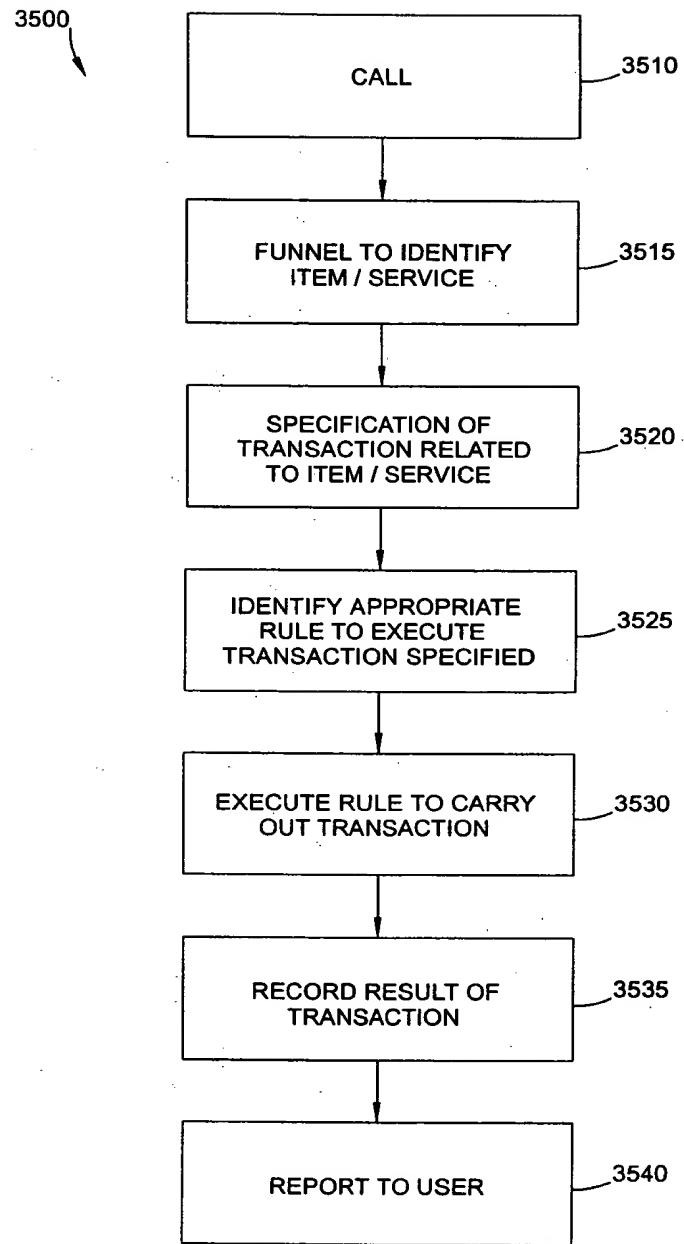


FIG. 35

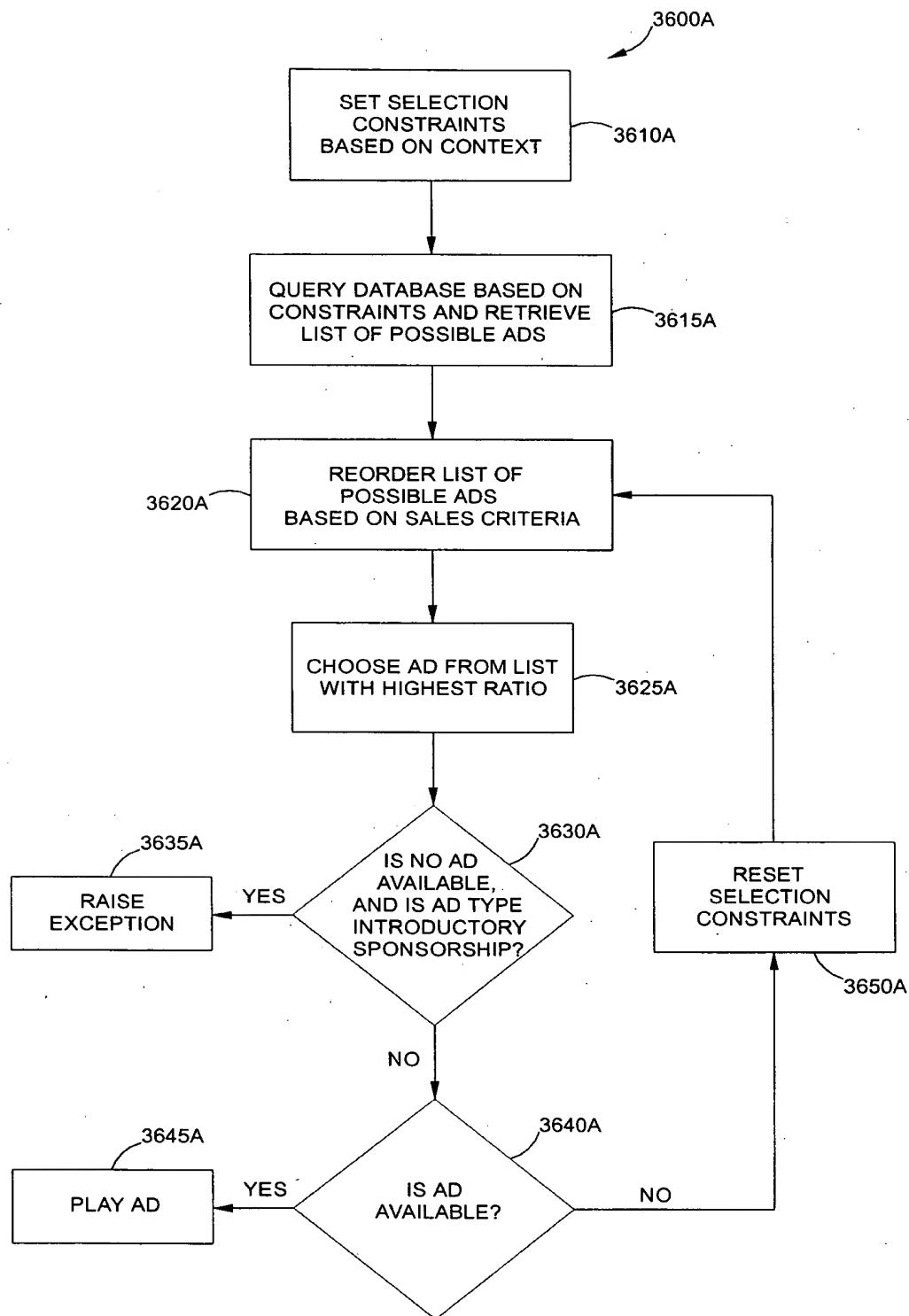


FIG. 36A

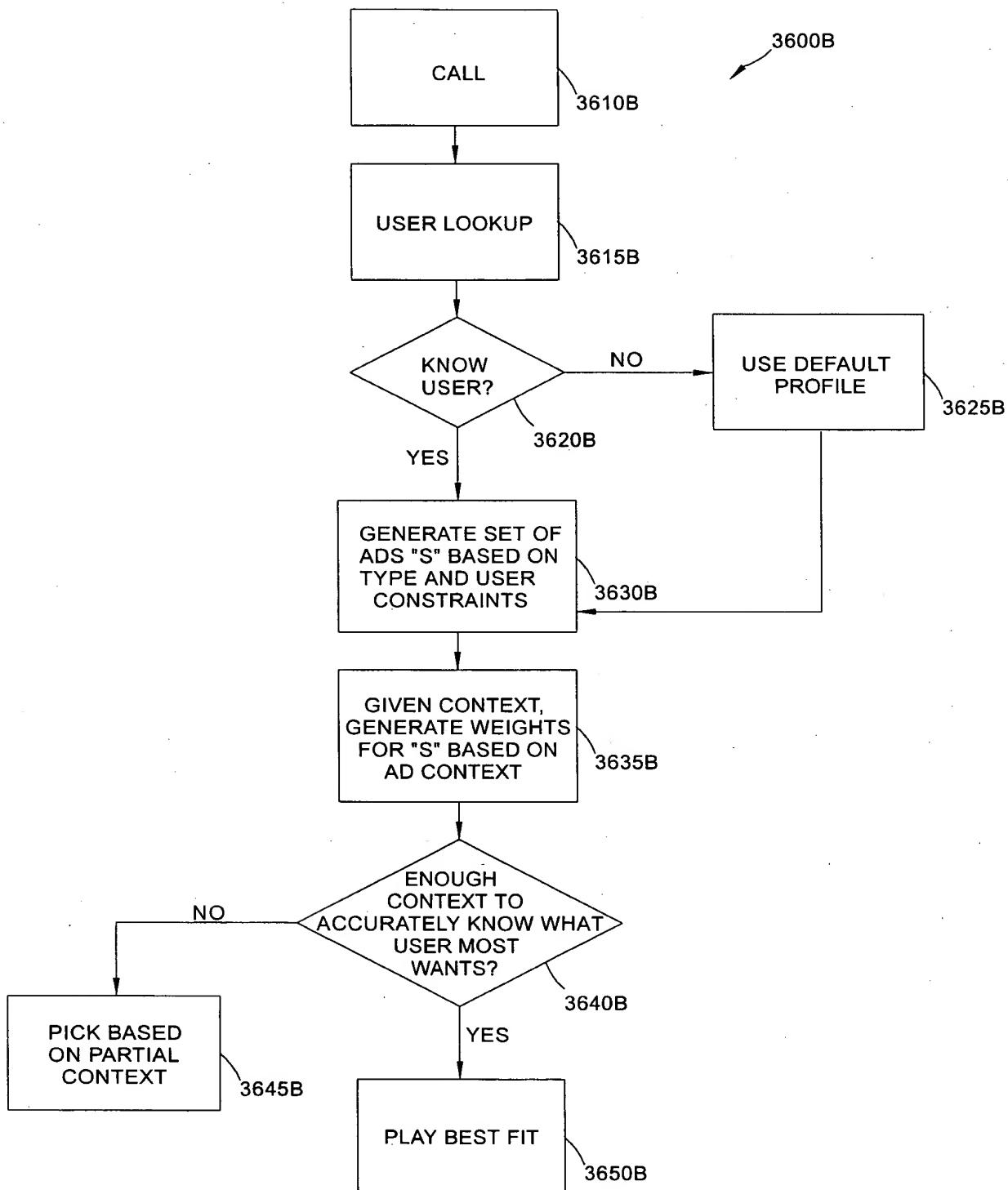
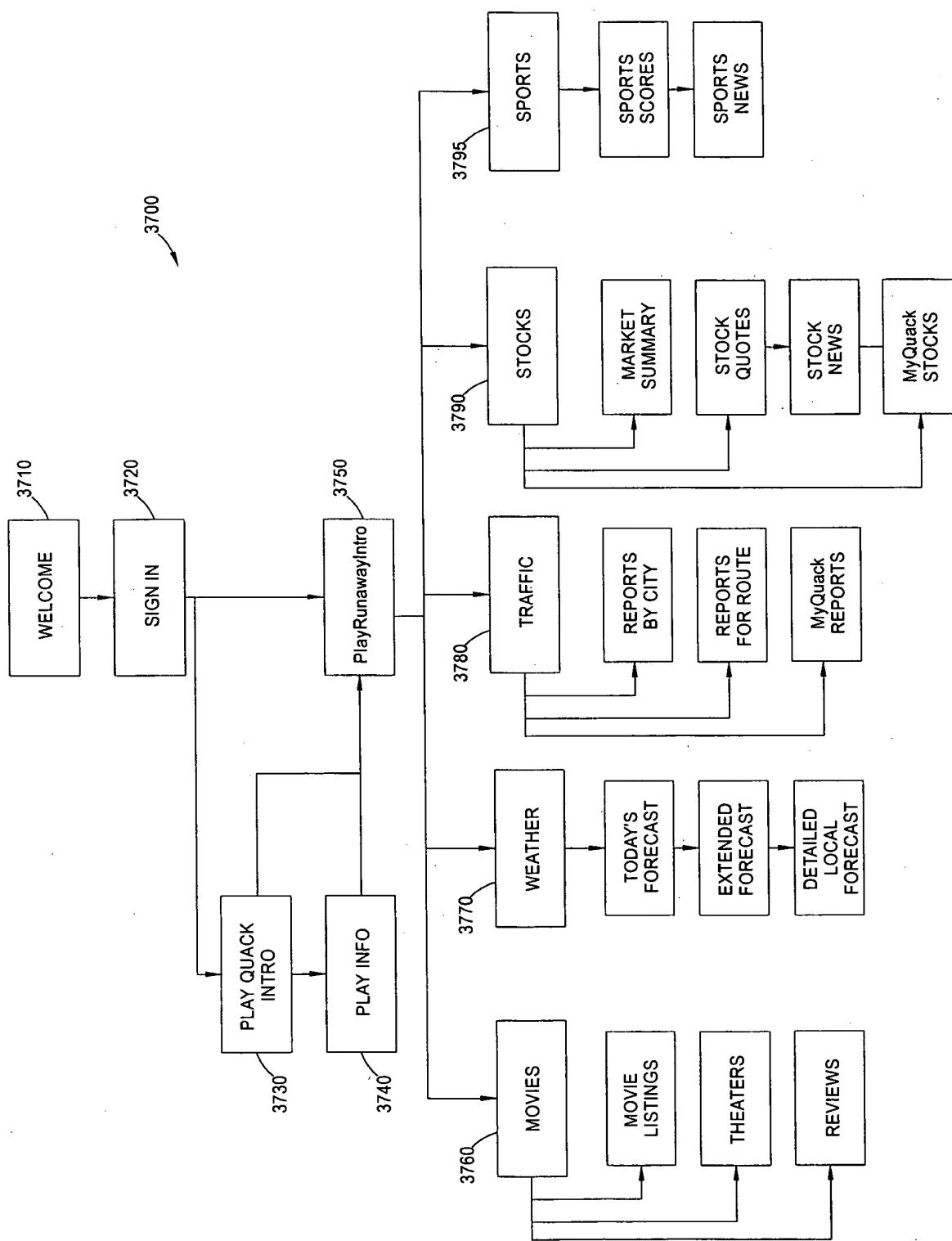


FIG. 36B

FIG. 37



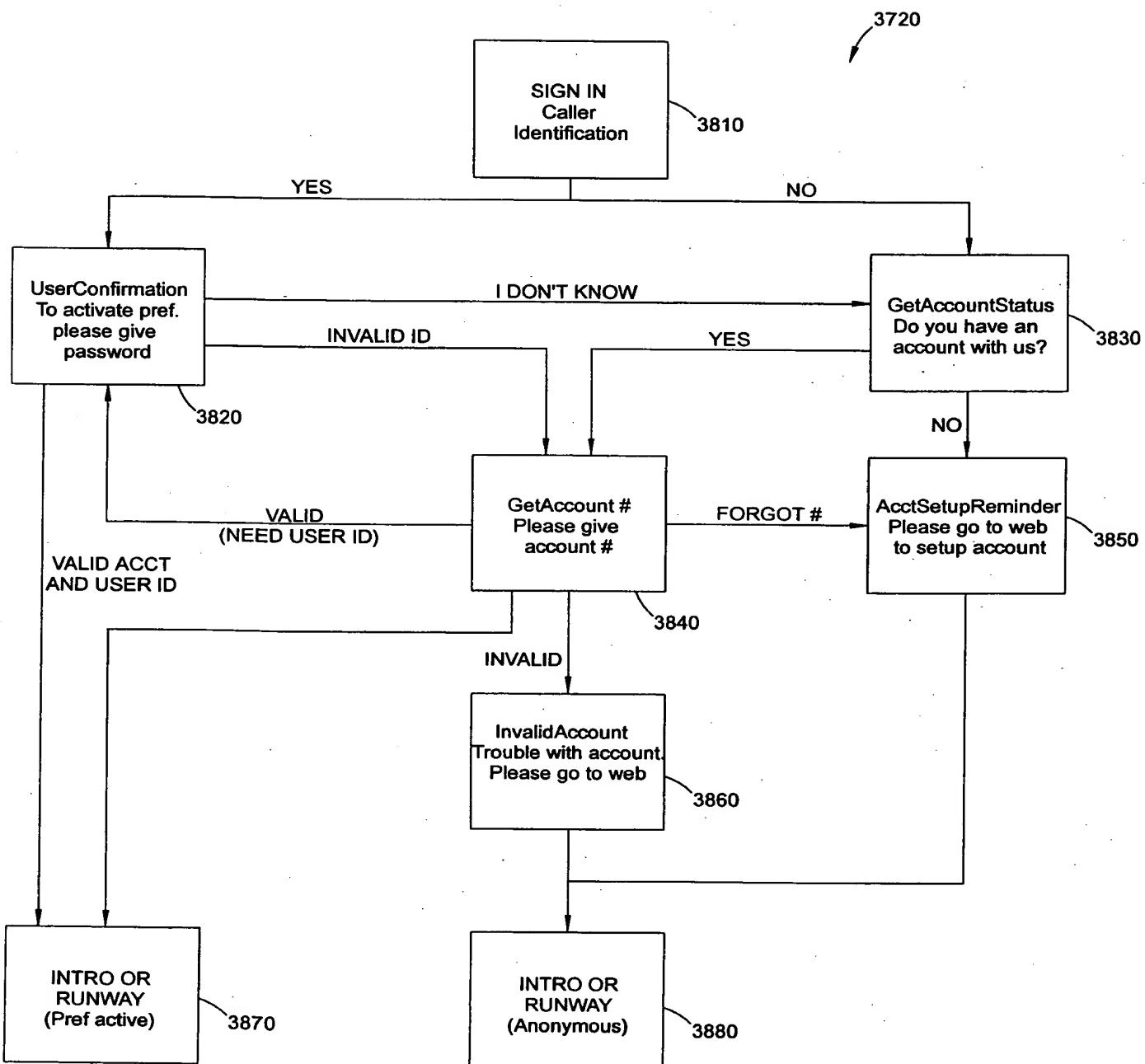


FIG. 38

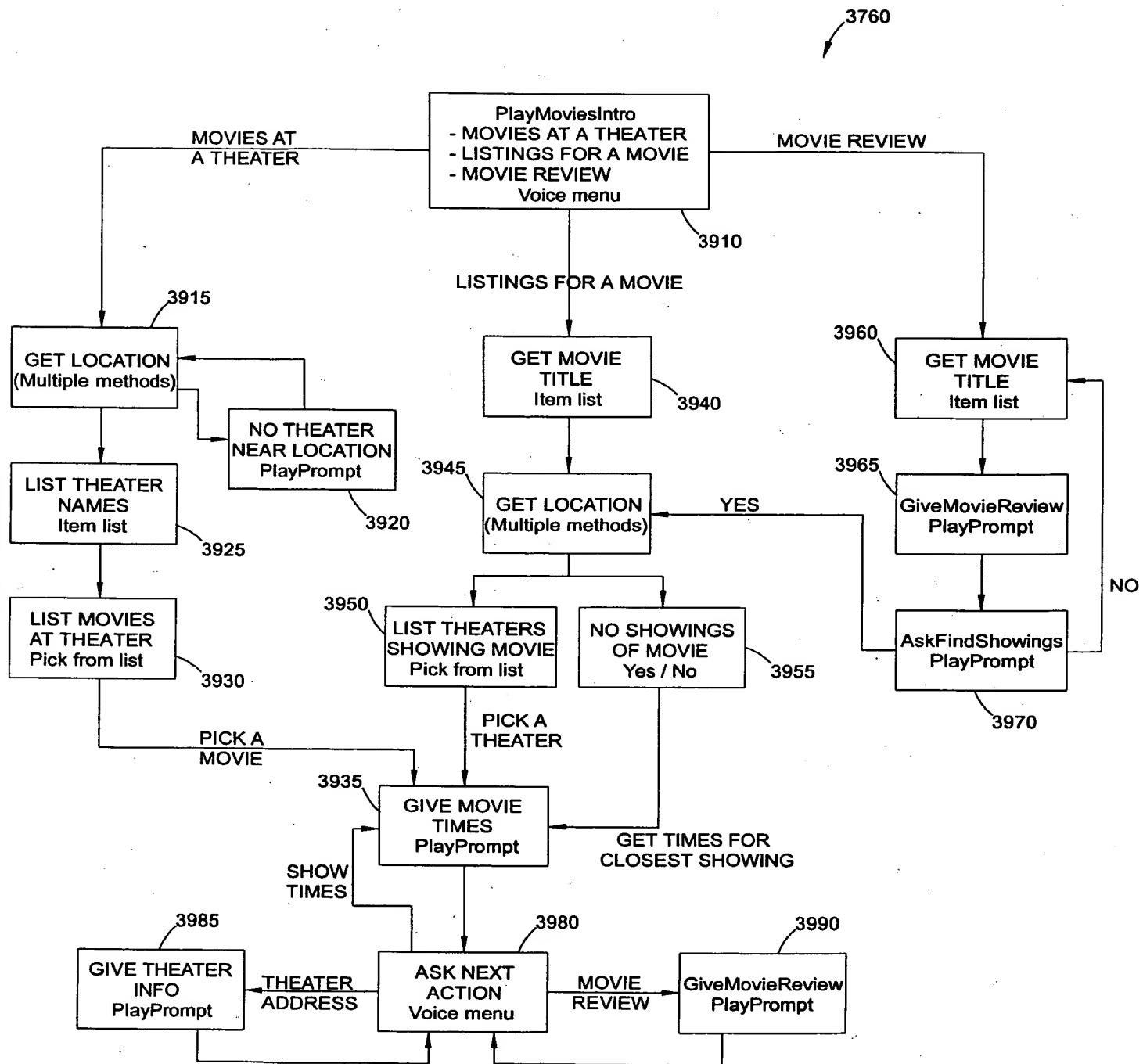


FIG. 39

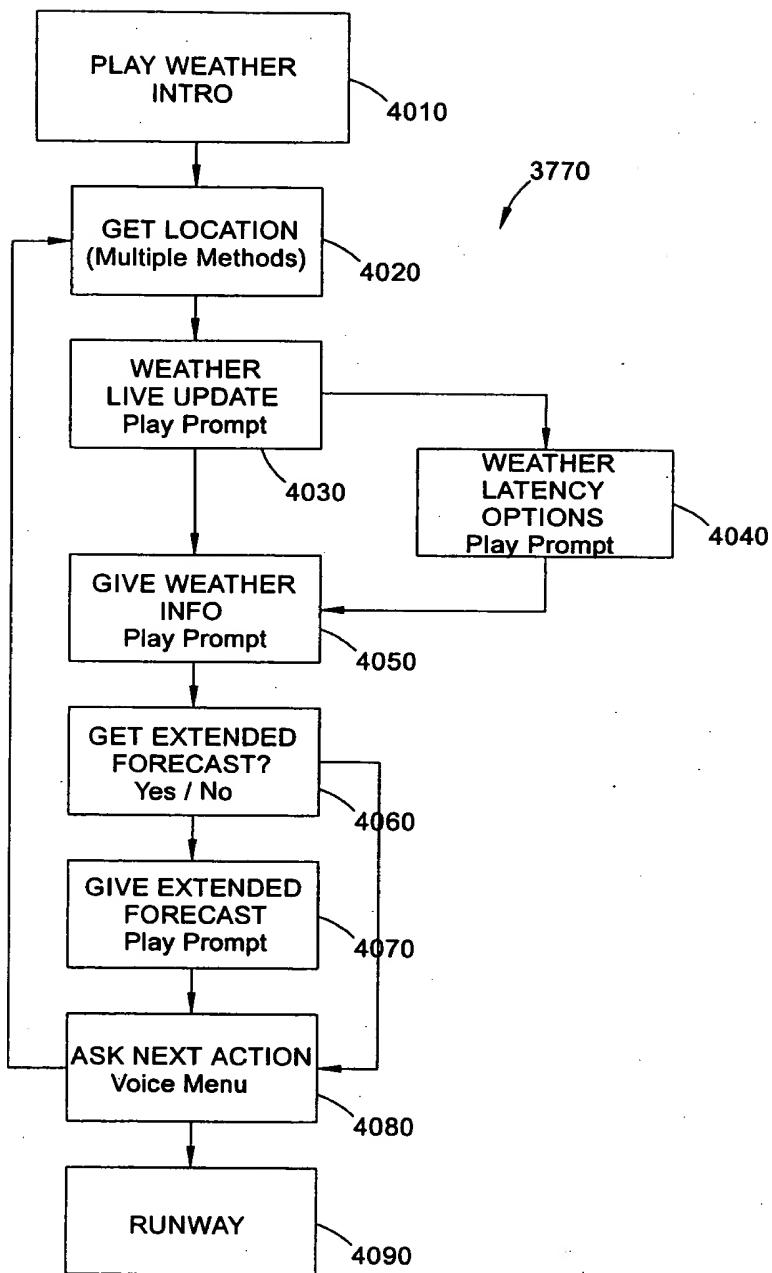


FIG. 40

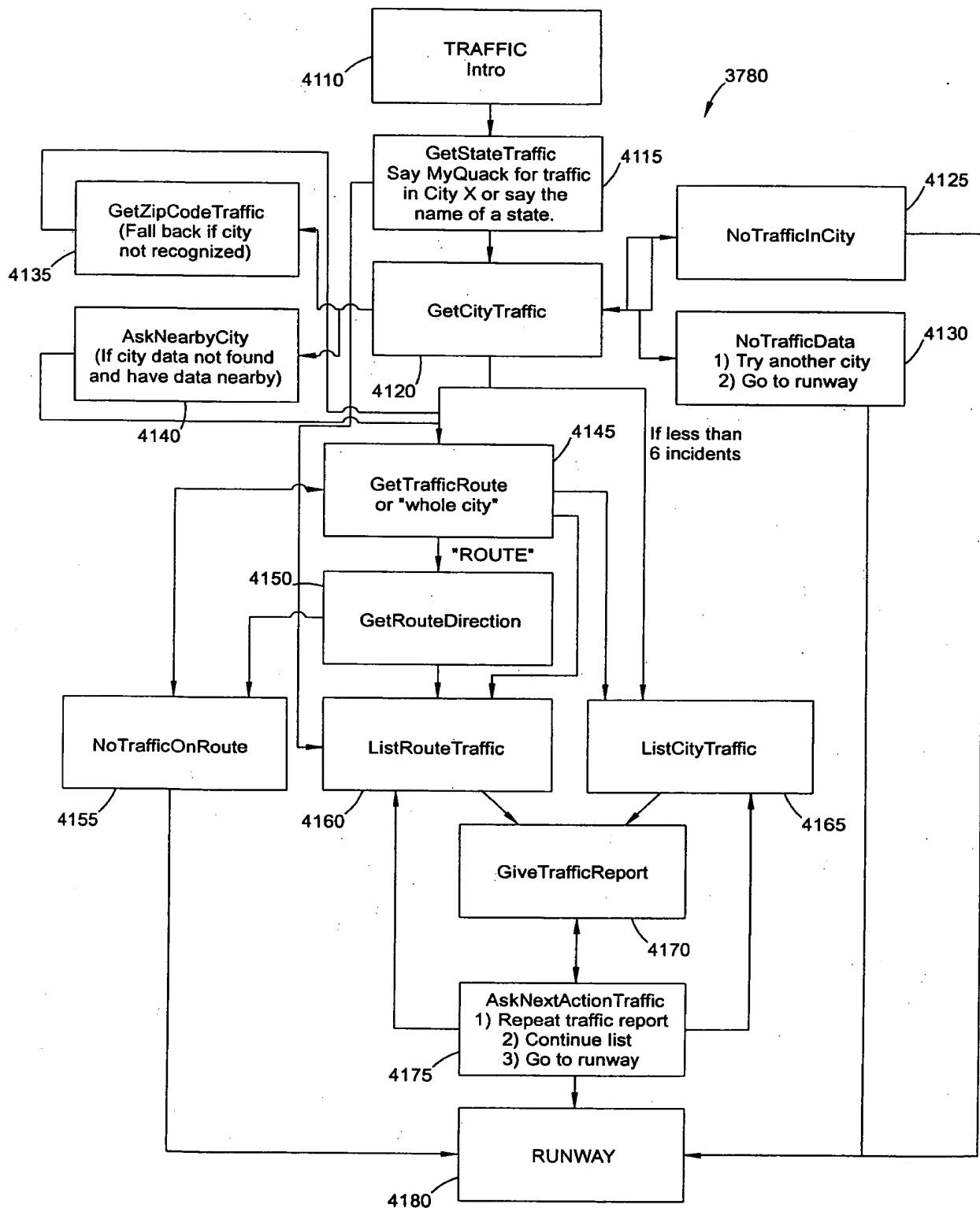


FIG. 41

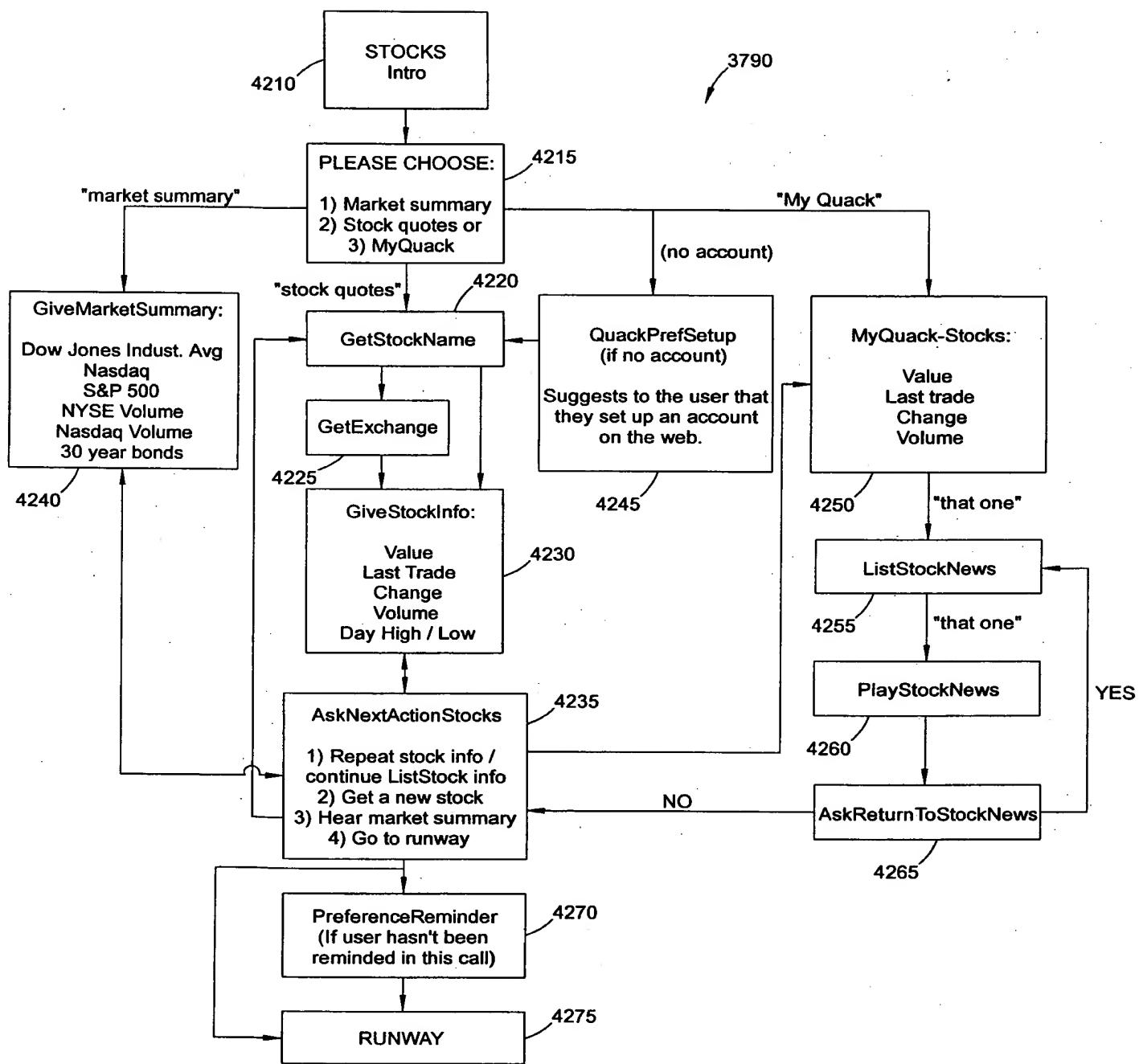


FIG. 42

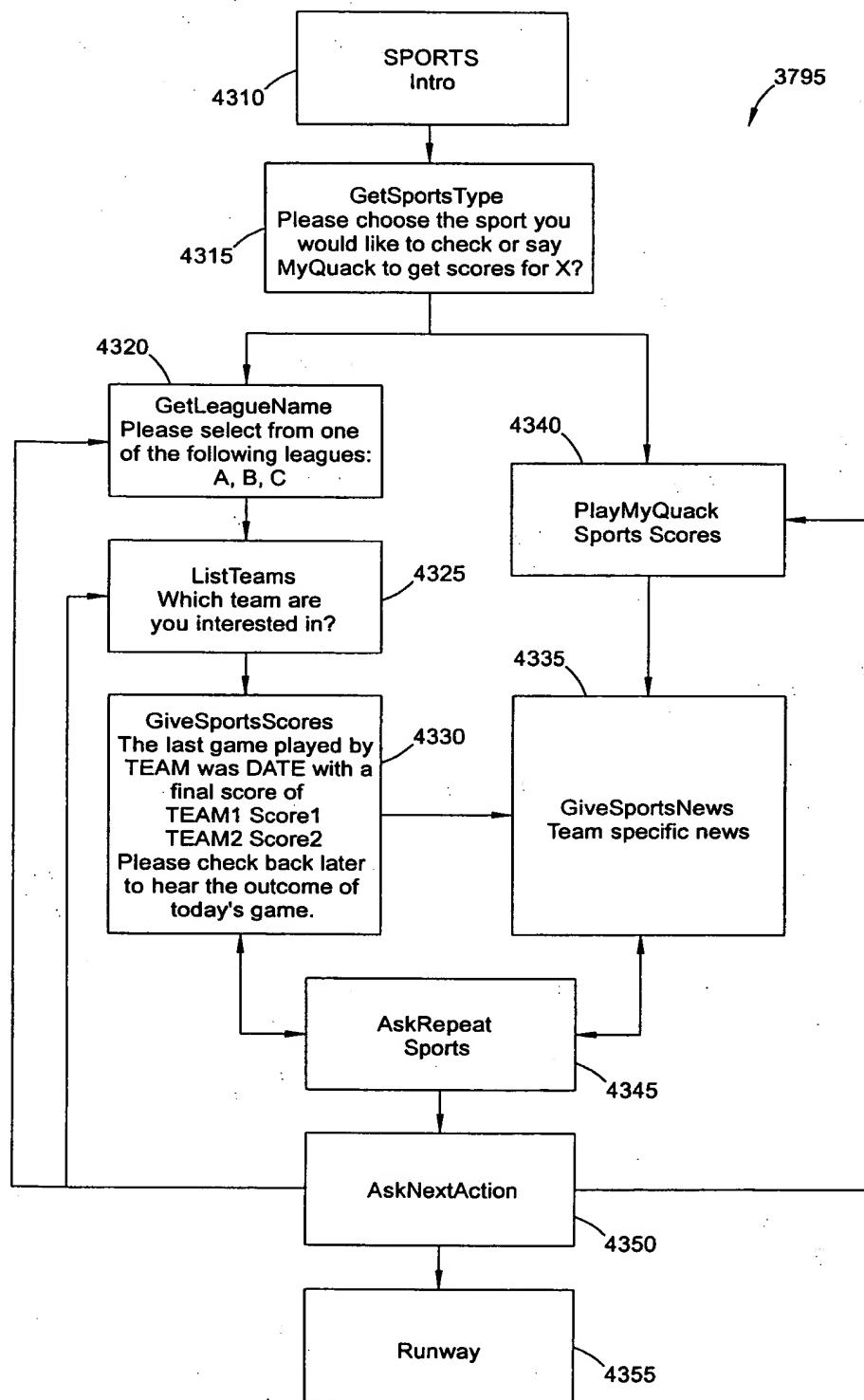


FIG. 43